Ekonomická fakulta
Faculty of Economics

Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

Conference Proceedings

37th International Conference on Mathematical Methods in Economics 2019

České Budějovice | September 11–13, 2019

Published by:
University of South Bohemia in České Budějovice, Faculty of Economics
Studentská 13, 370 05 České Budějovice, Czech Republic

Editors: Michal Houda, Radim Remeš

The publication has not passed the language correction.

ISBN 978-80-7394-760-6

Foreword

The international conference Mathematical Methods in Economics (MME) is a traditional annual meeting of professionals from universities and business interested in the theory and applications of operations research and econometrics. During the years, the conference has been recognized as a premier event in the field, bringing together researchers and practitioners in the industrial and institutional sectors sharing recent theoretical and applied results.

This proceedings is devoted to contributions presented at the 37th conference organized by the University of South Bohemia in České Budějovice (the Czech Republic), under the auspice of the Czech Society for Operations Research, the Slovak Society for Operations Research, and the Czech Econometric Society. We would like to thank the authors for their submissions, and more than 50 referees and programme committee members for their evaluations and valuable comments. The papers cover various topics such as operations research and decision making, data envelopment analysis, fuzzy modeling, or various mathematical, statistical, and econometric topics including financial, network, transportation, or project management models and applications.

The conference is organized by the Faculty of Economics, established in 2007 but holding the long tradition of economic university education in České Budějovice developed since the 1960's. The Faculty of Economics represents a significant educational and research institution which establishes an independent and creative environment supporting the knowledge of economic disciplines, the creation of innovations, and transfer and capitalisation of knowledge based on the mutual cooperation of teachers, students and representatives of the business sphere and public institutions from South Bohemia, the Czech Republic and from abroad. We are honoured to welcome the participants of the conference in such inspirative place and wish the pleasant stay in České Budějovice during the conference on Mathematical Methods in Economics 2019!

In České Budějovice, September 2019 Michal Houda

Programme Committee

doc. Dr. Ing. Miroslav Plevný
President of the Czech Society for Operations Research
University of West Bohemia, Faculty of Economics

prof. Ing. Josef Jablonský, CSc.
Vice-president of the Czech Society for Operations Research
University of Economics in Prague, Faculty of Informatics and Statistics

doc. Ing. Vladislav Bína, Ph.D.
University of Economics in Prague, Faculty of Management

prof. RNDr. Helena Brožová, CSc.
University of Life Science in Prague, Faculty of Economics and Management

prof. Ing. Mgr. Martin Dlouhý, Dr., MSc.
University of Economics in Prague, Faculty of Informatics and Statistics

doc. Ing. Jan Fábry, Ph.D.
University of Economics in Prague, Faculty of Informatics and Statistics

prof. RNDr. Ing. Petr Fiala, CSc., MBA
University of Economics in Prague, Faculty of Informatics and Statistics

Ing. Ludvík Friebel, Ph.D.
University of South Bohemia, Faculty of Economics

prof. Ing. Jana Hančlová, CSc.
Technical University of Ostrava, Faculty of Economics

prof. Radim Jiroušek, DrSc.
University of Economics in Prague, Faculty of Management

doc. RNDr. Jana Klicnarová, Ph.D.
University of South Bohemia, Faculty of Economics

doc. RNDr. Ing. Miloš Kopa, Ph.D.
Charles University, Faculty of Mathematics and Physics

doc. RNDr. Ing. Ladislav Lukáš, CSc.
University of West Bohemia, Faculty of Economics

Ing. Michal Novák, Ph.D.
University of Economics in Prague, Faculty of Management

Local Organizing Committee

Michal Houda (chair)
Jana Klicnarová
Nikola Soukupová
Milada Holoubková
Klára Vocetková
Ludvík Friebel
Richard Mikitin
Martina Matějčková

Contents

Application of the Cox regression model with time dependent parameters to unemployment data

Petr Volf[1]

Abstract. The contribution deals with the application of statistical analysis of the process of events, with the intensity described by a generalized version of Cox regression model. Namely, we study a case where the impact of covariates changes in time. Therefore the model with time dependent parameters should be applied. A method of model components non-parametric estimation is recalled, the flexibility of result is assessed with a goodness-of-fit test based on martingale residuals. The application concerns to the real data representing the job opportunities development and reduction, during a given period. They record the period of changes characterized by the increased employees fluctuation and staff reduction. Hence, the risk of leaving the company is changing. In particular, the risk of older persons increases, while the fluctuation concerns more the people with shorter time spent with the company. Both these covariates are considered and their impact to the risk analyzed.

Keywords: statistical survival analysis, unemployment study, Cox regression model, time-dependent parameters, goodness-of-fit.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

The paper studies the problem of flexible modelling of process of events in the framework of statistical survival analysis. As a rule, the model for events occurrence is based on the notion of counting process (registering, i.e. counting observed events) and its intensity given by corresponding model of the hazard rate. The running time could be the calendar time as well as the individual time of the object after its entrance to the study. The hazard rate can, moreover, depend on a set of covariates influencing the risk of events. Even these covariates can change their values during the time, the basic assumption states that at time t the values of covariates are known (are observable) up to t. Thus, they can also depend on the history of running counting process of events (by so called "endogenous" covariates). The most popular model of hazard rate in survival analysis is the Cox regression model based on the assumption of multiplicative influence of covariates to the hazard rate. Naturally, the model has a number of variants and generalizations. One direction leads to a more flexible regression part, generalizing its linear form to a nonlinear or to fully non-parametric. Another generalization, which is used here, corresponds to possible change of covariates impact in time and utilizes time-varying regression parameters (c.f. Murphy and Sen, 1991).

Naturally, a set of models for hazard rate is much richer, from parametric ones, for instance based on the Weibull distribution with its scale parameter being a function of covariates, up to fully nonparametric models (cf. the idea of "doubly cumulative hazard rate" in McKeague and Utikal, 1991). Another often used option is the additive Aalen model consisting of a sum of components, each representing the influence of one covariate. The "accelerated failure time" (AFT) model is popular as well, based on the assumption that the influence of covariate can accelerate (or slow down, on the contrary) the flow of individual object's time, i.e. it speeds up or slows down its growing, ageing, degeneration etc. Approaches and methods of event-data modeling are described in a number of papers and monographs, e.g. in Andersen et al (1993), Kalbfleisch and Prentice (2002).

The outline of the paper is the following: The next section introduces the Cox model variant with time-dependent parameters and the method of its analysis. Namely, a method of its non-parametric estimation is presented. Then, in part 3, the data are presented and their analysis provided. The application analyzes the process of job opportunities development and reduction, during a given period, in a company. The example is taken from Kadane and Woodworth (2004, see also "Data case K"), as their data are interesting by recording, after an initial interval of stable growth, the period of changes characterized by the increased employees fluctuation and staff reduction. Two covariates are available, namely the age of employees and the time spent with the company. Both are developing in time, too. It is due the non-stable conditions in the company that a standard Cox model does not describe the process sufficiently. The impact of covariates to the risk of leaving the company is changing during the period of observation. That is why the model with time-dependent regression parameters has to be utilized. In

[1]Department of Stochastic Informatics, ÚTIA AV ČR, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic, volf@utia.cas.cz

particular, the risk of older persons increases (which could be taken as a feature of discrimination), while the fluctuation concerns more the people with shorter time with company. Finally, in part 4, a method of a goodness-of-fit test based on martingale residuals is described and the flexibility of model is assessed with its aid.

2 Cox model with time-dependent parameters

Suppose that $i = 1, 2, ..., n$ objects are examined, each during time period $[0, t_i]$, where the final moment t_i is either the time of observed event (count) or the time of censoring. Random censoring from the right side is considered, as a rule. Let $N_i(t)$ denote the counting process of $i-th$ object, i.e. then $N_i(0) = 0$ and jumps to 1 at t_i, provided the event is observed. This is denoted by the indicator $\delta_i = 1$, while in the case of censoring by $\delta_i = 0$. Further, let $Y_i(t)$ on $t \geq 0$ denote the process indicating whether the object i is at t observed (then $Y_i(t) = 1$) or not ($Y_i(t) = 0$). Standardly $Y_i(t) = 1$ on $[0, t_i]$. Finally, denote by $N(t) = \sum_{i=1}^{n} N_i(t)$ the process counting events observed on all subjects. One of basic assumption is that the censoring mechanism does not depend on times of events and does not contain any information on the count times distribution (the censoring is independent and uninformative).

The behavior of $N_i(t)$ is governed by its intensity process, $\lambda_i(t) = h_i(t) \cdot Y_i(t)$, where $h_i(t)$ is the hazard rate of the event occurrence for subject i at time t. The Cox model specifies its form as

$$h_i(t) = h_0(t) \cdot \exp(\boldsymbol{\beta}' \boldsymbol{x}_i(t)), \tag{1}$$

where $\boldsymbol{x}_i(t) = (x_{i,1}(t), ..., x_{i,K}(t)'$, are values of K covariates (they may depend on time), $\boldsymbol{\beta}$ corresponding K-variate vector of regression parameters, and $h_0(t)$ is a common baseline hazard rate.

Standardly, the estimation of regression parameters is performed in the generalized maximal likelihood framework and is also described elsewhere (c.f. again Andersen et all, 1993, Kalbfleisch and Prentice, 2002). The full likelihood function can be written as

$$L = \prod_{i=1}^{n} h_i(t_i)^{\delta_i} \cdot \exp\{-\int_0^{t_i} h_i(t) dt\},$$

where $h_i(t)$ from (1) includes both unknown parameters as well as unspecified baseline hazard function. Estimates of regression parameters are obtained by the maximization of corresponding partial likelihood

$$L_p = \prod_{i=1}^{n} \Big(\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_i(t_i))}{\sum_{k=1}^{n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_k(t_i)) Y_k(t_i)} \Big)^{\delta_i}.$$

Then, the Breslow-Crowley estimator of increments of cumulative baseline hazard function $H_0(t) = \int_0^t h_0(s)\, ds$ is

$$\Delta \hat{H}_0(t) = \frac{dN(t)}{\sum_{k=1}^{n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_k(t)) Y_k(t)},$$

which is nonzero just at times of observed events, i.e. at $t = t_i$ with $\delta_i = 1$.

However, let us assume that the data indicate that the impact of covariates is changing during observation period and therefore the time-dependent parameters, i.e. functions $\boldsymbol{\beta} = \boldsymbol{\beta}(t)$, should be considered. It opens a question of their flexible estimation. The problem is solved quite similarly as in the standard regression model case: Either the functions are approximated by certain functional types (polynomial, combination of basic functions, regression splines) or constructed by a smoothing method, similar to moving window or kernel regression approach. The method adopted in Murphy and Sen (1991) is of such a type. Another method is based on the Bayes approach and treats coefficients $\boldsymbol{\beta}(t)$ as a random dynamic sequence with Gauss prior model of its development.

3 Real data example

The data are taken from the Statlib database: http://lib.stat.cmu.edu/datasets/caseK.txt, the "Case K" data. The data contain the records of all persons employed by a firm during the period of observation, from 1.1.1900 to 31.1.1995, namely their dates of birth, dates when persons were hired by the company and when they have left it, either voluntarily or were forced to leave (dismissed). There were together 412 people, from them 96 were fired, 108 left voluntarily, the rest, 208 employees, were still with the company at the end of data collection period. The time considered is the calendar time, in days, from 1 to 1857, the end of study is also the fixed time of censoring, namely $C = 1857$ is the upper bound for each personal record (it is so called type I censoring by fixed value). The number of employees at the study beginning was 163, it means that 249 people have joined the company during the followed period, i.e. their history in company started at certain $t_{0i} > 0$, thus changing the risk set of the study. Hence corresponding processes $Y_i(t) = 0$ on $(0, t_{0i})$.

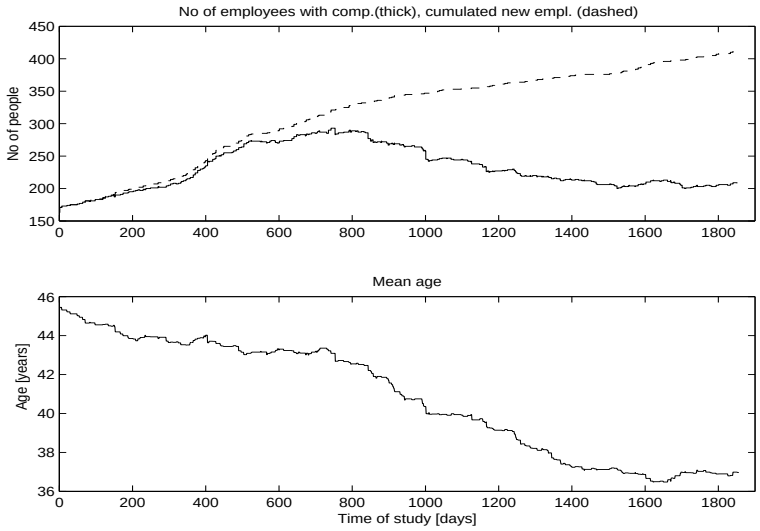


Figure 1 Development of number of employees during the study period (above). Thick curve shows actual number of employees, dashed displays the increase of newly hired people, i.e. the difference means the number of people who already left. The lower plot shows the development of average age of employees.

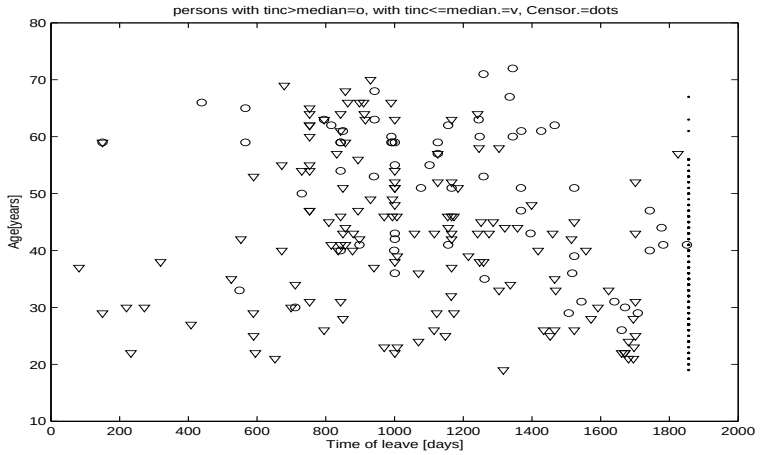


Figure 2 Times when employees have left the company vers. their age; circles for $tinc >$ its median, triangles for $tinc \leq$ median, dots denote censored records for people staying with company at the study termination.

The age of employees and the time spent with company (named here *"tinc"*) were taken as covariates, both were time-dependent. It was expected that both can influence the risk of leaving the company. The age varied from 20 to 70 years, its median was 39, while the time spent with the company varied from only 11 days to more than 41 years (15086 days), with median 1322 days. The changes in the company can be traced already from Figure 1, upper plot. The period of higher intensity of employees fluctuation (which could be taken as an indicator of certain non-stable conditions in the firm) started at about day 800 and lasted almost another 800 days. Figure 2 then shows graphically times of leaves, their dependence on age and also on the time spent with company $(tinc)$.

The analysis of these data appeared also in Kadane and Woodworth (2004). They used the Cox model and Bayes method mentioned above to analyze time changes of the impact of age to employees forced dismissals, the aim of their study was to explore whether older employees were discriminated having higher rate of dismissal. To do it, they explored the probability of regression parameter (of hazard of dismissals on age) being significantly positive, in Bayes sense. However, they did not take into account possible inter-dependence between the risk of forced and voluntary terminations. Such a dependence was studied in Volf (2018) in the framework of competing risks model, the mutual positive dependence was proved, simultaneously decreasing with the age. The interpretation was that the risk of being fired might lead the employees to decision to leave the company voluntarily, in time, and that such a preference concerned more the younger people, thus decreasing the occurrence of forced dismissals for them. However, it was due computational complexity that in Volf (2018) just constant Cox parameters and fully parametric model were considered. The present study does not distinguish between the ways of employment

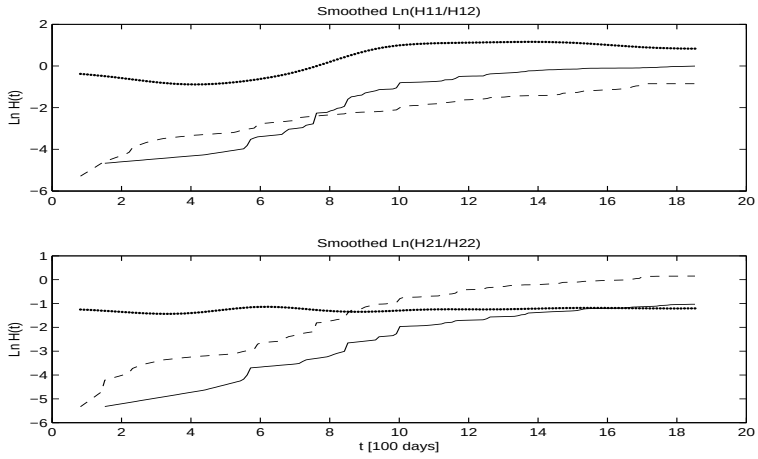


Figure 3 Above: Logarithms of cumulated hazard rates for persons with age$>$ 39 years (full curve) and age$\leq$ 39 (dashed), and their smoothed difference (thick). Below, similarly for $tinc > 1322$ days (full), for $tinc \leq 1322$ (dashed), and their smoothed difference (thick).

termination, on the other hand, the regression parameters are allowed to vary in time, thus making the model rather general.

3.1 Results

Figure 3 presents preliminary graphical test of whether the impact of covariates to the risk of leave the company can be modelled by the proportional hazard model. Namely, we observe difference of logarithms of plain cumulated hazard rates estimated from data with age grater or smaller than 39 years (upper graph) and from data with time with company greater or smaller than 1322 days (both being the medians of covariates). While in the lower plot the curves are approximately parallel, the upper plot shows intersecting curves, thus indicating that the impact of age was changing during the observation period. Hence, we decided to use the Cox model variant with time-dependent "parameters" $\beta_k(t)$, $k = 1, 2$, for both covariates, though Figure 3 does not indicate its necessity in the case of dependence on the time with company. Initial rough estimates of $\beta_k(t)$ were obtained via the moving window method (a variant of the method of sieves described e.g. in Murphy and Sen, 1991). Then, this rough estimates were smoothed with the aid of a kernel smoothing. Both estimates are displayed in Figure 4. It is seen that also $\beta_2(t)$ shows interesting variability and decrease in the second part of observation period. Prevailingly positive β_1 means that the risk of leaving the company increased with larger age, simultaneously $\beta_2 < 0$ indicates higher fluctuation among people with shorter time spent in the company.

Finally, Figure 5 shows the estimate of corresponding cumulated baseline hazard rate. It is seen that at the right end the baseline CHR increases sharply, which is the consequence of the fact that the impact of age decreased in the end part of observation period. It is seen both from Figure 2 and from the shape of estimated $\beta_1(t)$.

4 Goodness-of-fit tests

The selection and evaluation of a reasonable model is just one step of statistical analysis. The goodness-of-fit tests should follow. In the framework of intensity models for lifetime data, the goodness-of-fit tests are often based on the analysis of residual process (martingale residuals). The residual process is defined as the difference between estimated cumulative intensity and observed counting process of failures (see for instance Andersen et al, 1993), formally

$$R(t) = \hat{A}(t) - N(t) = \hat{A}(t) - A(t) - M(t),$$

where $M(t)$ is a martingale, $A(t) = \int_0^t \sum_{i=1}^n h_i(s)Y_i(s)ds$ is the cumulative intensity process (simultaneously it is the variance process of $M(t)$) and $\hat{A}(t) = \int_0^t \sum_{i=1}^n \hat{h_i}(s)Y_i(s)ds$ is its estimate. In the framework of Cox model version (1), $\hat{h_i}(s)ds \sim \Delta\hat{H}_0(s) \cdot \exp(\hat{\boldsymbol{\beta}}'(s)\,\boldsymbol{x}_i(s))$ at counts points $s = t_i$.

Hence, the residual process is constructed from observed data, its properties depend on properties of estimators. In the case without regression, as well as in the Aalen additive regression model, residual processes are the martingales, too (Volf, 1996). In some other cases, as are the Cox model or the accelerated failure time model, the behavior of estimates, and therefore of residuals, is more complicated. Notice also that in the Cox model framework the sum over all subjects yields $\hat{A}(t) = N(t)$ directly, that is why the residuals are as a rule formulated more

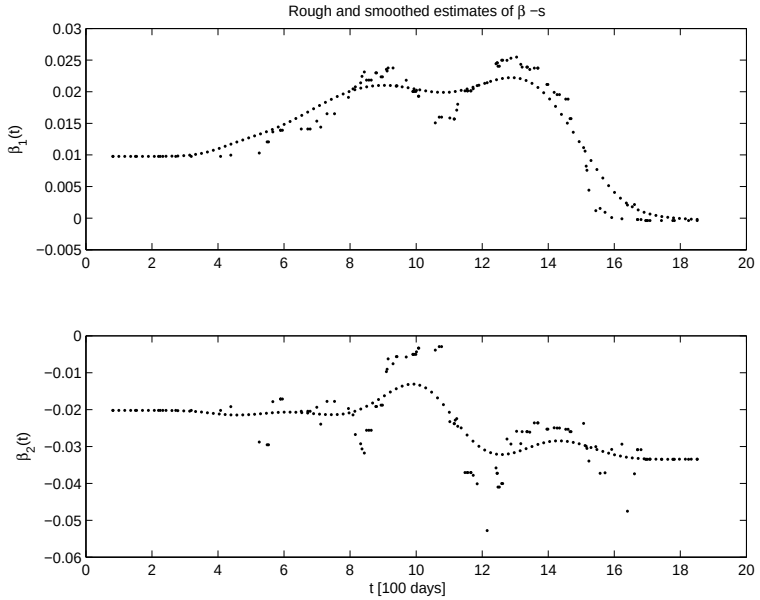


Figure 4 $\beta_{1,2}(t)$ estimated by a mowing window method, at data time points (scattered dots), then their kernel-smoothed versions (smooth curves).

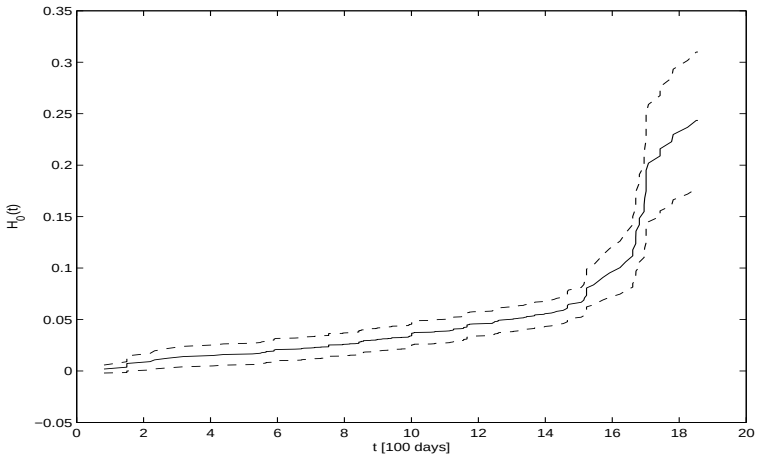


Figure 5 Estimated cumulated baseline hazard rate, with point-wise 95% confidence intervals

generally, namely

$$dR(t) = \sum_{i=1}^{n} W_i(t) \cdot (dN_i(t) - d\hat{A}_i(t)),$$

with some convenient weight processes $W_i(t)$. The simplest choice takes $W_i(t) = 1[i \in S]$, i.e. the indicator of a set of indices (a stratum) $S \subset \{1, 2, ..., n\}$. Then stratified residuals are obtained.

As it has been already said, in the Cox model case the residual processes are not the martingales, though, after a proper standardization, they asymptotically tend to Gauss processes. Therefore, a random generation of would-be residual processes under the hypothesis of model fit is possible. By such a generation we obtain a sample of 'ideal' residual processes. Then, certain characteristics of generated residuals are compared with the same characteristics obtained from the data. Anyway, practical tests of Cox model fit are often performed just graphically, comparing visually how far are residuals in group S from zero line, or, equivalently, $\hat{A}_S(t) = \sum_{i \in S} \hat{A}_i(t)$ from $N_S(t) \sum_{i \in S} N_i(t)$, as in Arjas (1988). We shall use the same method in the present case.

In order to check the fit of proposed model, we stratified the data randomly, many times, selecting a subgroup, strata S, and displaying $R_S(t) = N_S(t) - \hat{A}_S(t)$. It is seen that the residual processes oscillate around zero level, thus assessing a good fit of the model. As the residual processes should, approximately, represent a Gauss processes with zero means and growing variances given by cumulated intensities $A_S(t)$, that is why also the variability of

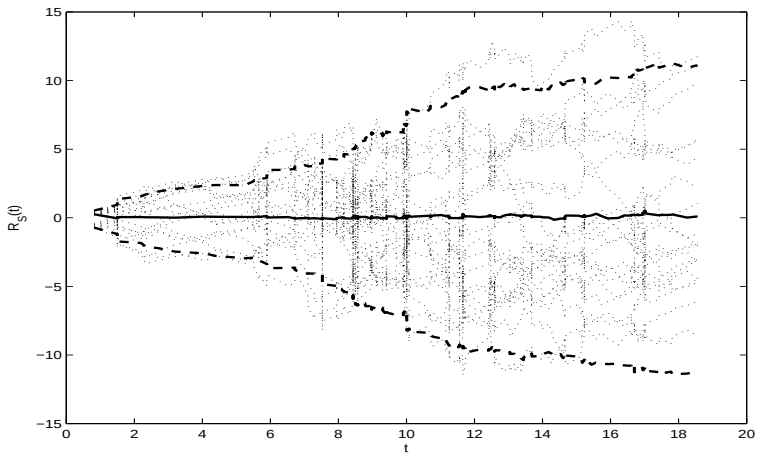


Figure 6 Graphical goodness of fit test: Empirical point-wise median and 5%, 95% quantiles from residual processes computed in 1000 randomly selected data sub-samples. Some residual processes are visualized (dotted).

processes in Figure 6 increases with growing time.

5 Concluding remarks

The contribution had in fact two goals. The first purpose was to present a real case of count data with regression such that the impact of regressors to the risk of count is changing in time. Then also the model had to reflect this phenomenon. Hence, the selection of model and presentation of methodology of its evaluation was the second purpose of the paper. From a set of convenient models the Cox model with time-dependent parameters was chosen, as the Cox model ranks among the most popular ones in the field of statistical survival analysis, also due its reasonable interpretation. The model suitability was confirmed by an appropriate goodness-of-fit test.

Acknowledgement: The research was supported by the grant No. 18-02739S of the Grant Agency of the Czech Republic.

References

[1] Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazard model. *J. Amer. Statist. Assoc.,* 83, 204–212.

[2] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer.

[3] Kadane, J.B., Woodworth, G.G. (2004). Hierarchical Models for Employment Decisions. *Journal of Business and Economic Statistics*, 22, 182–193.

[4] Kalbfleisch, J.D., Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

[5] McKeague, I. W., Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand.J. Statist.*, 18, 177-195.

[6] Murphy, S.A., Sen, P.K. (1991). Time-dependent coeffisients in a Cox-type regression model. *Stoch. proc. and Applications,* 39, 153–180.

[7] Volf, P. (1996). Analysis of generalized residuals in hazard regression models. *Kybernetika,* 32, 501–510.

[8] Volf, P. (2018). Problem of competing risks with covariates: Application to an unemployment study. In *Proceedings of the 36th International Conference Mathematical Methods in Economics,* FM VŠE Jindřichův Hradec, 624–629.

[9] Data case K: Downloaded 04.03.2019 from *http://lib.stat.cmu.edu/datasets/caseK.txt*

VRP with loading time window

Jan Pelikan[1]

Abstract. The vehicle routing problem (VRP) is a standard task solving distribution of goods from the center to customers. There are a number of modifications, such as VRPs with time windows, VRPs with the external carrier, VRP with heterogeneous fleet etc. This contribution contains a new VRP modification, a dynamic delivery problem, i.e. the problem of delivering goods within a certain time interval. Customers time window relates to the time for loading customers demand. The time of loading the customers demand on a vehicle and therefore the time of the vehicle to exit the depot and start the route to customers must be within the time interval (time window) of all these customers. Travel time from the depot to customer and the unloading time are not considered and do not effect to the optimal solution. A mathematical model is presented and a savings heuristic is modified for this problem. An illustrative example is attached.

Keywords: dynamic vehicle routing problem, savings heuristic, integer programming.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Similarly to the classic vehicle routing problem (VRP) alternatively capacitated VRP, it is a task on undirect complete graph, where the nodes represent customers. One node is a depot where both vehicles and goods are located. Let n be the number of nodes, where node 1 is a depot and C is a matrix of distances between nodes, where c_{ij} is the distance between node i and node j. Node j, where $j > 1$, is customers node with the demand q_j and the time interval for its loading is $[\underline{t_j}, \bar{t}_j]$. Time is discrete, therefore the vehicles can be loaded only at time points from the set $I = \{1, 2, \ldots, T\}$.

Customers time window relates to the time for loading customers demand. The time of loading the customers demand on a vehicle and therefore the time of the vehicle to exit the depot and start the route to customers must be within the time interval (time window) of all these customers. Travel time from the depot to customer and the unloading time are not considered and do not effect to the optimal solution. The capacity of vehicles W cannot be exceeded. The goal is to minimize the total length of the routes of all vehicles. Therefore the vehicle and time of loading its demand have to be assigned to each node . At the same time, it is necessary to minimize the total length of the routes of all vehicles. The demand of the node i must be loaded in the time interval $[\underline{t_i}, \bar{t}_i]$ at this point, the vehicle will be leaving the depot for its route to customers. For example, in case studies may be given the day of the week when the demand of node is available for loading into vehicle and the number of days during those the demand has to be loaded for transport. During these days the vehicle must start shipping the demand.

It is obvious that the classical VRP can be reduced to this VRP modification called vehicle routing problem with time windows for loading (VRPTWL), which makes this modification also NP hard. When considering a greater number of nodes, the mathematical model will not be able to solve it within a reasonable time, so heuristic methods will be useful in these cases. For this task, the saving method is modified in this article.

This VRP modification resembles vehicle routing problem with time window (VRPTW), but it has a different conditions. For VRPTW, each node is given a time interval during which the vehicle must visit this node and hand over the goods. A time when the vehicle visits the node depends on the time when it leaves the previous node, the route and the transport time from the previous node. Loading and

[1]University of Economics, W. Churchill sq. 4, Prague, pelikan@vse.cz

unloading time of the goods also can be added to that time. The order of the nodes on the route then ensures that the vehicle visits the node in the given time interval, in the time window. The VRP problem is described in the literature, heuristic methods are proposed, and a summary of these approaches is described in [1], [2], [3].

The proposed problem is different from the classic VRP with time windows at the nodes and it can not be converted into this problem. Initially, a mathematical model of the task will be proposed followed by savings heuristic modifications.

2 Mathematical model of the problem

The mathematical model VRPTWL is based on the VRP model, where the limiting factor is the vehicle capacity, but the number of vehicles is not limited. This capacity cannot be exceeded when creating vehicle routes. Therefore the sum of demands of all nodes on the route must not exceed this capacity. Each route will be realized at a certain time from the set $I = \{1, 2, \ldots, T\}$, and this time point must be included in the time windows of all nodes on the route. Therefore, it is necessary to determine the time when the vehicle is loaded with the demands of the nodes on the route and assures that this time is within the intervals $[\underline{t_i}, \bar{t}_i]$ of all nodes laying on the route.

Parameters of the model are:
n number of nodes,
$c_{i,j}$ distance between node i and node j,
q_i demand of node i,
W capacity of vehicle,
$\bar{t}_i$, the latest time of loading the demand i,
$\underline{t_i}$ the earliest time of loading the demand i.

Variables of the model are:
x_{ij}^t binary, equals 1 if a vehicle loading in time t travels from node i to node j,
u_i^t variables in anti-cyclic constraints.

Mathematical model of VRPTWL:

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}^t \to min \tag{1}$$

$$\sum_{i=1}^{n} x_{ij}^t = \sum_{i=1}^{n} x_{ji}^t, \quad j = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T, \tag{2}$$

$$\sum_{j=1}^{n}\sum_{t=1}^{T} x_{ij}^t = 1, \quad i = 2, 3, \ldots, n, \tag{3}$$

$$\sum_{j=1}^{n}\sum_{t=1}^{\underline{t_i}-1} x_{ij}^t = 0, \quad i = 1, 2, \ldots, n, \tag{4}$$

$$\sum_{j=1}^{n}\sum_{t=\bar{t_i}+1}^{T} x_{ij}^t = 0, \quad i = 1, 2, \ldots, n, \tag{5}$$

$$u_i^t + q_j - W(1 - x_{ij}^t) \leq u_j^t, \quad t = 1, 2, \ldots, T \quad i, j = 1 \ldots, n \quad i \neq j \quad j > 1, \tag{6}$$

$$u_i^t \leq W, \quad i = 1, 2, \ldots, n \quad t = 1, 2, \ldots, T \tag{7}$$

$$x_{ij}^t \text{binary}, \quad u_i^t \geq 0, \quad i = 1, \ldots, n \quad t = 1, 2, \ldots, T. \tag{8}$$

In this formulation, the objective function (1) minimizes the sum of distances of all vehicles routes. Constraint (2) states that if the vehicle loading in time t enters the node j must leave this node in time t. Equation (3) means that one vehicle in one time enters node i and assurs it's demand q_i. Constraints (4) and (5) ensure that the supply for the node i, i.e. the loading of the demand of node i, must take

place in the time interval $[\underline{t_i}, \overline{t}_i]$. Anti-cyclic conditions are in (6). Inequality (7) assures that capacity of vehicles are not exceeded. The proposed mathematical model was verified on an illustrative example. The results are shown in the next section.

3 Savings heuristic

Savings heuristic for VRP with time windows for loading is a modification of the savings heuristic created for the original VRP (see [3]). It is a savings heuristic, where the routes are gradually merged into a new route. The default route system consists of simple routes containing a depot and one node that represents a customer.

Notation.
$\mathcal{R}$ is a set of routes, $\mathcal{R}$= $\{R^s\}$, where R^s is s-th route with load $w(R^s)$ and the distance of the route $d(R^s)$.

Step 1: { initial routes} :

Saving numbers $s_{ij} = c_{1,i} + c_{j,1} - c_{i,j}$ are calculated, the set of saving numbers is denoted S. Lets create a initial set of routes $\mathcal{R} = \{R^s\}$ where $R^s = \{(1, s), (s, 1)\}, s = 2, 3, \ldots, n$, the set of nodes of this route except for depot is $V^s = \{s\}$.

The length of the route is $d(R^s) = 2c_{1,s}$, the size of the freight of the route R^s is denoted $w(R^s) = q_s$, the time interval of this route is $tw(R^s) = [\underline{t_s}, \overline{t}_s], s = 2, 3, \ldots, n$.

Step 2: {two routes R^a and R^b choice} :

If the set S is empty the heuristic ends, otherwise we will choose maximal advantage number $s_{k,l}$ and put $S := S - \{s_{k,l}\}$.

Lets search for two routes R^a and R^b, $a \neq b$, from $\mathcal{R}$ such that following conditions have to meet:

- C1: edge $(k, 1)$ is from R^a and edge $(1, l)$ is from R^b,
- C2: $w(R^a) + w(R^b) <= W$,
- C3: $tw(R^a) \cap tw(R^b)$ is not $\emptyset$.

If there is no pair of routes meet conditions C1, C2, C3 then we repeat Step 2

Step 3: {merging R^a and R^b} :

- $R^c := R^a \cup R^b \cup \{(k, l)\} - \{(k, 1), (1, l)\}$,
- $w(R^c) := w(R^a) + w(R^b)$,
- $d(R^c) := d(R^a) + d(R^b) - (c_{k,1} + c_{1,l}) + c_{k,l}$,
- $tw(R^c) := tw(R^a) \cap tw(R^b)$,
- $\mathcal{R} := \mathcal{R} \cup R^c - (R^a \cup R^b))$,
- go to Step 2.

4 Numerical example

Consider 11 nodes, node 1 is a depot, the capacity of each vehicle is W=100. The node requirements and delivery intervals are in Table 1, distance matrix C is:

$$C = \begin{pmatrix} 0 & 13 & 6 & 55 & 93 & 164 & 166 & 168 & 169 & 241 & 212 \\ 13 & 0 & 11 & 66 & 110 & 175 & 177 & 179 & 180 & 239 & 208 \\ 6 & 11 & 0 & 60 & 97 & 168 & 171 & 173 & 174 & 239 & 209 \\ 55 & 66 & 60 & 0 & 82 & 113 & 115 & 117 & 117 & 295 & 265 \\ 93 & 261 & 97 & 82 & 0 & 113 & 115 & 117 & 118 & 333 & 302 \\ 164 & 175 & 168 & 113 & 113 & 0 & 6 & 4 & 2 & 403 & 374 \\ 166 & 177 & 171 & 115 & 115 & 6 & 0 & 8 & 7 & 406 & 376 \\ 168 & 179 & 173 & 117 & 117 & 4 & 8 & 0 & 2 & 408 & 378 \\ 169 & 180 & 174 & 117 & 118 & 2 & 7 & 2 & 0 & 409 & 379 \\ 241 & 239 & 239 & 295 & 333 & 403 & 406 & 408 & 409 & 0 & 46 \\ 212 & 208 & 209 & 265 & 302 & 374 & 376 & 378 & 379 & 46 & 0 \end{pmatrix}$$

node i	1	2	3	4	5	6	7	8	9	10	11
q_i	0	19	24	30	20	35	25	32	20	22	37
$[\underline{t}_i, \bar{t}_i]$	[1,7]	[2,4]	[1,3]	[3,7]	[4,7]	[2,6]	[5,6]	[4,6]	[3,6]	[2,6]	[5,7]

Table 1 example

We create a initial set of routes $\mathcal{R} := \{R^s\}, R^s = \{(1,s),(s,1)\}, s = 2,3\ldots,n$ and time interval and transport volume for each route. We also calculate and sort out the downwardly saving numbers s_{ij}. Route merging according to a descending sequence of saving numbers is next. Value for each saving number will be skipped if it breaks the rule for connecting routes, otherwise, a linked route will be created:

- $s_{10,11}$=407 we will link the route 1-10-1 a 1-11-1 to create route 1-10-11-1 transport volume is 57 and time interval [5,6],
- $s_{8,9}$=335 we will link the route 1-8-1 a 1-9-1 to create route 1-8-9-1, transport volume is 52, time interval [4,6],
- $s_{6,9}$=331 we will link the route 1-8-9-1 a 1-6-1 to create route 1-8-9-6-1, transport volume is 87, time interval [4,6],
- $s_{7,9}$=328 does not meet the condition C1, $s_{6,8}$=328 does not meet the condition C1, $s_{7,8}$=326 does not meet the condition C2, $s_{6,7}$=324 does not meet the condition C2,
- $s_{5,7}$=144 we will link the route 1-5-1 a 1-7-1 to create route 1-5-7-1 transport volume is 45 and time interval [3,6],
- $s_{5,6}$=144 does not meet the condition C2, $s_{5,8}$=144 does not meet the condition C2, $s_{5,9}$=144 does not meet the condition C1, $s_{4,9}$=107 does not meet the condition C1, $s_{4,6}$=106 does not meet the condition C2,
- $s_{4,7}$=106 we will link the route 1-5-7-1 a 1-4-1 to create route 1-5-7-4-1 transport volume is 75 and time interval [3,6],
- $s_{4,8}$=106 does not meet the condition C2, $s_{4,5}$=66 does not meet the condition C1, $s_{2,11}$=17 does not meet the condition C3, $s_{2,10}$=15 does not meet the condition C3,
- $s_{2,3}$=8 we will link the route 1-2-1 a 1-3-1 to create route 1-2-3-1 transport volume is 48 and time interval [2,3],
- For other advantageous numbers, no routes can be linked without breaking C1 or C2 or C3.

The heuristic method results in the following routes:

1. route 1-2-3-1 with the interval [2,3] , transport volume 48 and length of the route 30,
2. route 1-8-9-6-1 with the interval [4,6], transport volume 87 and length of the route 336,

3. route 1-5-7-4-1 with the interval [3,6], transport volume 75 and length of the route 378,

4. route 1-10-11-1 with the interval [5,6], transport volume 57 and length of the route 499,

The total length of all routes is 1243. If we solve the task using the mathematical model, it can be shown that this solution is optimal. Each route can be realized at any time from the specified interval. At this time it is necessary to load vehicles with requirements of the nodes from this route.

Conclusion

An interesting modification for vehicle routing problem is studied, a mathematical model of the problem is presented and the new heuristic method is proposed. An illustrative example is added.

References

[1] Laporte G. : The vehicle routing problem: an overview of exact and approximate algorithms. *European Journal for Operations Research* **59** (1992), 345-358.

[2] Desrochers M., Desrosiers J., Solomon M.,M., : A new optimalization algorithm for the vehicle routing problem with time windows. In: *Operations Research* **40** (1992), 342-354.

[3] Braysy O., Gendreau M. : the vehicle routing problem with time windows, part I. route construction and local search algorithms. *Transportation Science* **39** (2005), 104-118.

Implicitly weighted robust estimation of quantiles in linear regression

Jan Kalina[1], Petra Vidnerová[2]

Abstract. Estimation of quantiles represents a very important task in econometric regression modeling, while the standard regression quantiles machinery is well developed as well as popular with a large number of econometric applications. Although regression quantiles are commonly known as robust tools, they are vulnerable to the presence of leverage points in the data. We propose here a novel approach for the linear regression based on a specific version of the least weighted squares estimator, together with an additional estimator based only on observations between two different novel quantiles. The new methods are conceptually simple and comprehensible. Without the ambition to derive theoretical properties of the novel methods, numerical computations reveal them to perform comparably to standard regression quantiles, if the data are not contaminated by outliers. Moreover, the new methods seem much more robust on a simulated dataset with severe leverage points.

Keywords: regression quantiles, robust regression, outliers, leverage points.

JEL classification: C14
AMS classification: 62G35

1 Introduction

In a variety of linear regression applications, it is useful and/or necessary to estimate quantiles instead of only the mean (average) trend. There have been numerous applications of regression quantiles e.g. in finance, economics, medicine or engineering. Successful applications of regression quantiles were overviewed in a paper [13] or in a recent monograph [14] devoted entirely to quantile regression. In finance, popular quantile risk measures include expected shortfall or value at risk (VaR), where the latter is equal exactly to a quantile. In (nonlinear) time series, it is sometimes desirable to find a baseline (such as a baseline hazard) after detecting and smoothing peaks [3]. Regression quantiles are also commonly used as a tool for modeling economic growth [15]; in this context, it is perhaps the only common approach suitable for modeling growth curves with heteroscedastic errors.

The methodology of regression quantiles (also known as quantile regression), which was proposed back in 1990s and investigated in seminal works of [5, 12], represents now a popular concept in linear regression. In the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \ldots, n, \tag{1}$$

the regression τ-quantile for $\tau \in (0, 1)$ is defined as a (regression) line with parameters obtained as

$$\arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^T b \right), \tag{2}$$

where $X_i = (X_{i1}, \ldots, X_{ip})^T$ denotes the i-th observation and ρ_τ (defined in [12]) is considered in the form

$$\rho_\tau(x) = x \left(\tau - \mathbb{1}[x < 0] \right), \quad x \in \mathbb{R}, \tag{3}$$

with indicator function denoted by $\mathbb{1}$. Alternatively, ρ_τ may be formulated as

$$\rho_\tau(x) = \begin{cases} \tau x & \text{if } x \geq 0, \\ (\tau - 1)x & \text{if } x < 0. \end{cases} \tag{4}$$

In (1), the usual assumption $\mathsf{var}\, e_i = \sigma^2$ is denoted as homoscedasticity (for some fixed $\sigma > 0$), while heteroscedasticity is defined simply as its violation. The least squares estimator b_{LS} of β in (1) is not efficient and the

[1]The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Praha 8, Czech Republic, kalina@cs.cas.cz

[2]The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Praha 8, Czech Republic, petra@cs.cas.cz

corresponding estimate of var b_{LS} is biased under heteroscedasticity. On the other hand, regression quantiles remain reliable under heteroscedasticity and represent the most suitable tool for modeling heteroscedastic data [12]. They are able to capture the whole distribution of the response (i.e. conditioning on the regressors) instead of simply considering the mean trend. Thus, if the whole set of regression τ-quantiles is evaluated for various values $\tau \in (0, 1)$, heteroscedasticity may be easily revealed. At the same time, it is possible to use regression quantiles to test for heteroscedasticity.

In the analysis of real data, it remains important to use methods resistant against the presence of outliers or measurement errors [8, 2]. Regression quantiles are known to be robust (resistant) to values outlying on the vertical axis. However, they are not robust to leverage points [1], i.e. to values outlying on the horizontal axis. If the distribution of random errors in the regression model is not symmetric, then the standard regression quantiles are known to be inadequate and an alternative (more robust) appraoch was proposed in [4]. Another shortcoming of standard regression quantiles is the phenomenon known as quantile crossing, i.e. a low stability of regression quantiles across various values of τ. Because this has been reported as embarrassing [7], researchers have attempted to study alternative versions of regression quantiles. Crossing of regression quantiles may be interpreted as their instability, while a stable (robust) alternative remains highly desirable.

Estimating regression quantiles represents a very important task also in nonlinear regression. Therefore, the standard regression quantiles (2) have been extended to nonlinear regression in a straightforward way already in [16], where a known nonlinear function f in the model

$$Y_i = f\left(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}\right) + e_i, \quad i = 1, \ldots, n, \tag{5}$$

was considered. Other extensions are available for quantile regression with a multivariate response [17] or lasso estimation of regression quantiles proposed already in [12].

The aim of this paper is to propose an alternative approach to regression quantile estimation, based on the highly robust least weighted squares estimator. The novel implicitly weighted quantile estimator is proposed in Section 2. Section 3 considers robust estimation exploiting the novel quantiles. Section 4 illustrates the performance of the novel quantiles on real as well as simulated datasets. Finally, Section 5 concludes the paper.

2 Implicitly weighted quantiles for linear regression

In the whole paper, the standard linear regression model (1) is considered. The least weighted squares (LWS) estimator proposed by [19] represents a promising robust estimator of parameters of (1), which performs well in regression applications [9] or dimensionality reduction [11]. The aim of this section is to propose an implicitly weighted estimator of regression quantiles, conceptually similar to the LWS.

We need to use the concept of ranks for univariate random variables [6], defined in the Appendix. The new estimator will be defined by means of a weight function denoted as ψ, which is again defined in the Appendix, following the notation of [19]. Let us now use the notation $u_1(b), \ldots, u_n(b)$ for residuals corresponding to a fixed $b \in \mathbb{R}^p$. Their ranks (following Definition 2) will be denoted by $R_1(b), \ldots, R_n(b)$ to stress the dependence on b. Keeping in mind that both these vectors depend on b, we define the estimator as

$$\underset{b \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} \psi_1\left(\frac{R_i(b)}{n}\right) \left(u_i(b)\right)^2. \tag{6}$$

The novel estimator denoted as quantile least weighted squares (QLWS) will be now defined.

Definition 1 (QLWS estimator)**.** Assuming two given weight functions ψ_1 and ψ_2, the QLWS estimator is defined as

$$\underset{b \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} \left[\psi_1\left(\frac{R_i(b)}{n}\right) \mathbb{1}_{[u_i(b)>0]} + \psi_2\left(\frac{R_i(b)}{n}\right) \mathbb{1}_{[u_i(b)<0]}\right] \left(u_i(b)\right)^2. \tag{7}$$

Particularly, if it is chosen $\psi_2 = c\psi_1$ for a given $c > 0$, then (7) reduces to

$$\min \sum_{i=1}^{n} \psi_1\left(\frac{R_i(b)}{n}\right) \left(\mathbb{1}_{[u_i(b)>0]} + c\mathbb{1}_{[u_i(b)<0]}\right) \left(u_i(b)\right)^2. \tag{8}$$

The choice $c = 1$ corresponds to the LWS estimator itself, while $c > 1$ estimates a bottom quantile and $c > 1$ an upper quantile. Thus, c is a parameter similar to τ for standard regression quantiles (but with a different interpretation). It is recommendable that the user chooses several different values of c to examine the results for upper as well as lower quantiles. Concerning the choice of the weight function, we recommend to use weights

ensuring a high breakdown point, e.g. the trimmed linearly decreasing weights proposed in [10]. The QLWS estimator is based on implicit weighting and thus we consider (7) or (8) to be generalizations of the LWS (6). It may be conjectured that the QLWS retains (i.e. inherits) the high local robustness of the LWS, although its rigorous investigation extends the scope of this paper. The QLWS estimator is not based on any τ, but the user may compute the novel quantiles for various values of c above as well as below 1. The computation of the QLWS estimator may exploit directly the FAST-LTS algorithm of [18].

3 Robust regression by interquantile estimation

In linear regression (1), the trimmed least squares (TLS) computed from regression quantiles represents a popular estimator of β with appealing robustness properties [8]. Let us extend the idea to compute standard estimates between two novel quantiles.

Let us first recall that the TLS estimator in (1) which is well described in basic textbooks of robust statistics (e.g. [8]), is formally computed as the least squares estimator between two regression quantiles with selected values of τ. Particularly, the procedure for its computation starts with obtaining two regression quantiles, performs outlier deletion based on them, and finally estimates the trend by the least squares computed over all remaining observations. Under certain assumptions, the TLS estimator is asymptotically normal, regression equivariant, and (unlike most other robust regression estimators) also scale equivariant [8]. The TLS estimator has a bounded influence (and thus good local robustness properties, like the sample median), but a zero breakdown point, i.e. remains very vulnerable to severe outliers [8].

The novel estimator denoted as IQLWS (interquantile least weighted squares) is based on the QLWS of Section 2. Our novel definitions of estimators exploiting the principle of the TLS trim away observations below the lower quantile and above the upper quantile, ensuring robustness of the resulting estimate by deleting a given percentage of most outlying observations. In other words, the extreme observations obtain (and retain) zero weights and then the regression line is subsequently fitted by least squares. Considering the model (1), we use the notation IQLWS(c_1, c_2) for the interquantile estimator computed for two quantiles with $c = c_1$ and $c = c_2$.

4 Examples

The QLWS and IQLWS as novel methods require to be investigated thoroughly on a number of real as well as simulated datasets. The aim of the examples is to illustrate the performance of the novel methods for linear as well as nonlinear regression. Our aim is to reveal that the novel quantiles are suitable under heteroscedasticity, outliers, and particularly leverage points, where the last concepts is commonly used to denote outlying values on the horizontal axis (in values of one or more regressors). In the computations, we use the definition (8) for the QLWS, using trimmed linear weights proposed in [10].

4.1 Investment data

The first example is devoted to a real dataset of U.S. investment data previously analyzed in [9] and shown in Figure 1. Real gross private domestic investments are modeled here as a response of the real gross domestic product. Both variables are evaluated in the United States in the years from 1980 to 2001 in 10^9 of U.S. dollars. Originally the data come from the website `www.stls.frb.org/fred` of the U.S. Department of Commerce.

Figure 1 (left) shows standard regression quantiles computed for $\tau = 0.2$ (the bottom quantile) and $\tau = 0.8$ (the top quantile). Figure 1 (right) shows the results of the QLWS. Numerical values of different estimators for two regression parameters (intercept β_0 and slope β_1) are given in Table 1, where QLWS(1) is exactly equal to the LWS (with the same weight function). This dataset is simplistic without apparent outliers or leverage points. The two lines corresponding to the QLWS are more parallel (i.e. more stable) compared to the lines of standard regression quantiles. There are also results of the robust estimators TLS and IQLWS presented in Table 1.

4.2 Simulation A

In the first simulation, the dataset is shown in Figure 2. The data with a single regressor $X_1, \ldots, X_n$ are generated in a heteroscedastic linear regression model with $\beta_0 = 2$ and $\beta_1 = 1$ with normally distributed errors; the distribution of e_i for $i = 1, \ldots, n$ is $\mathsf{N}(0, \sigma_i^2)$ with $\sigma_i = X_i/2$. Estimated values of parameters are shown in Table 1 and visualized in Figure 2 for standard regression quantiles (left) as well as QLWS (right). The estimated trends of standard regression quantiles and QLWS are very similar and both methods seem equally able to work under the presence of heteroscedastic errors.

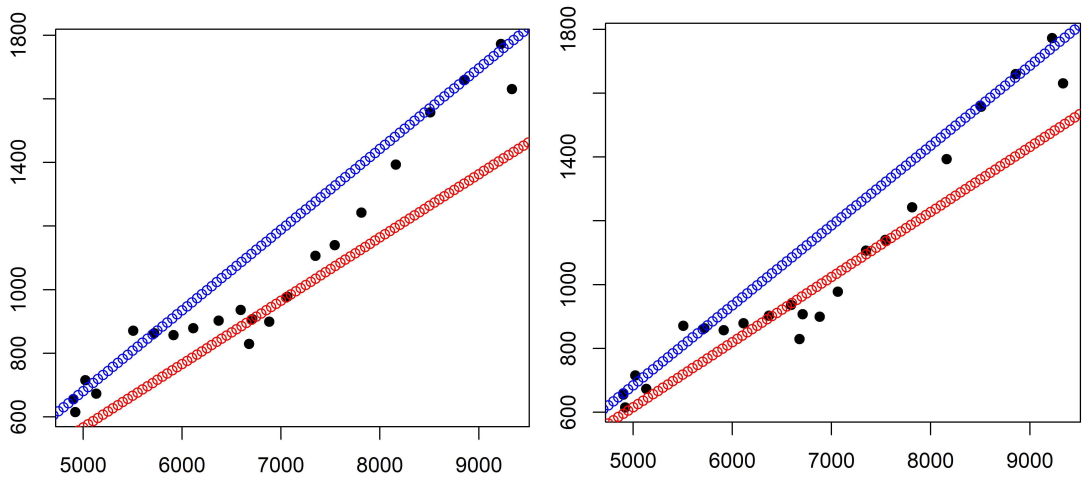


Figure 1 Investment data of Section 4.1. Left: results of standard regression quantiles with $\tau = 0.2$ (bottom) and $\tau = 0.8$ (top). Right: results of the QLWS estimator with $c = 0.005$ (top) and $c = 5$ (bottom).

Table 1 Estimated parameters of (standard) regression quantiles (RQ), QLWS, TLS and IQLWS for three different datasets.

	Investment		Simulation A		Simulation B	
Method	β_0	β_1	β_0	β_1	β_0	β_1
RQ(0.2)	-477	0.21	2.75	0.56	16.6	-0.28
RQ(0.5)	-516	0.23	2.31	0.99	8.5	0.36
RQ(0.8)	-587	0.25	2.50	1.34	10.9	0.41
QLWS(0.005)	-475	0.22	4.83	0.59	-2.0	0.37
QLWS(1) = LWS	-496	0.23	5.62	0.85	5.2	0.43
QLWS(5)	-568	0.25	7.76	1.32	7.8	0.49
TLS(0.2, 0.8)	-564	0.24	2.38	1.06	9.7	0.15
IQLWS(0.005, 5)	-503	0.23	5.51	0.78	4.7	0.42

4.3 Simulation B

Finally, a more difficult task is considered due to the presence of leverage points in the data. The simulated data are shown in Figure 3. The data are created as a mixture of 80 observations following a linear trend with a positive slope ($\beta_1 = 1/2$), where the errors are normal $\mathsf{N}(0, \sigma^2)$ with $\sigma = 25$, and 20 observations following a linear trend with a negative slope ($\beta_1 = -1/2$), where the errors are again normal but with $\sigma = 50$. Estimated values of parameters are shown again in Table 1 and visualized in Figure 3 for standard regression quantiles (left) as well as QLWS (right).

The results of the QLWS are very different from those of standard regression quantiles in this example. The lower of the standard regression quantiles corresponds to the contamination (i.e. focus very clearly on the minority trend with a negative slope), while the lower of the QLWS quantiles is close to being parallel with the upper quantile and with the majority trend with a positive slope. This reveals the robustness of the QLWS estimator.

5 Conclusions

Estimating regression quantiles represents an important task in a variety of economic applications. Standard regression quantiles have been very well theoretically investigated and there is a huge empirical experience with their practical performance over real econometric data. Still, regression quantiles are known to be heavily connected to assumptions of symmetric distribution of the random regression errors or no leverage points in the data. Nevertheless, leverage points are not uncommon in real econometric data.

Therefore, a novel estimator of quantiles in linear regression denoted as QLWS is proposed in this paper. It is an advantage that for the choice $c = 1$, the QLWS estimator is exactly equal to the LWS; at the same time, the

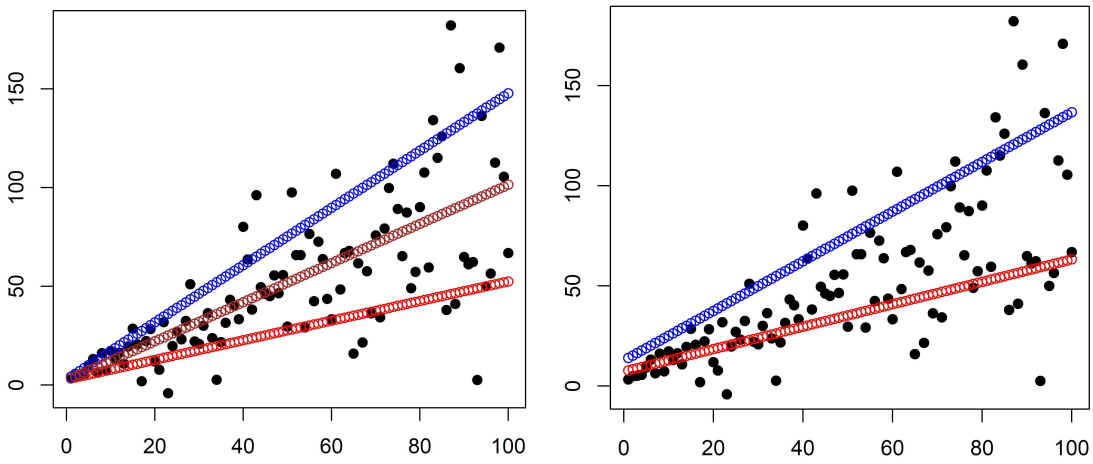


Figure 2 Simulation A of Section 4.2. Left: results of standard regression quantiles with $\tau = 0.1$ (bottom), $\tau = 0.5$ (middle) and $\tau = 0.9$ (top). Right: results of the QLWS estimator with $c = 0.005$ (top) and $c = 5$ (bottom).

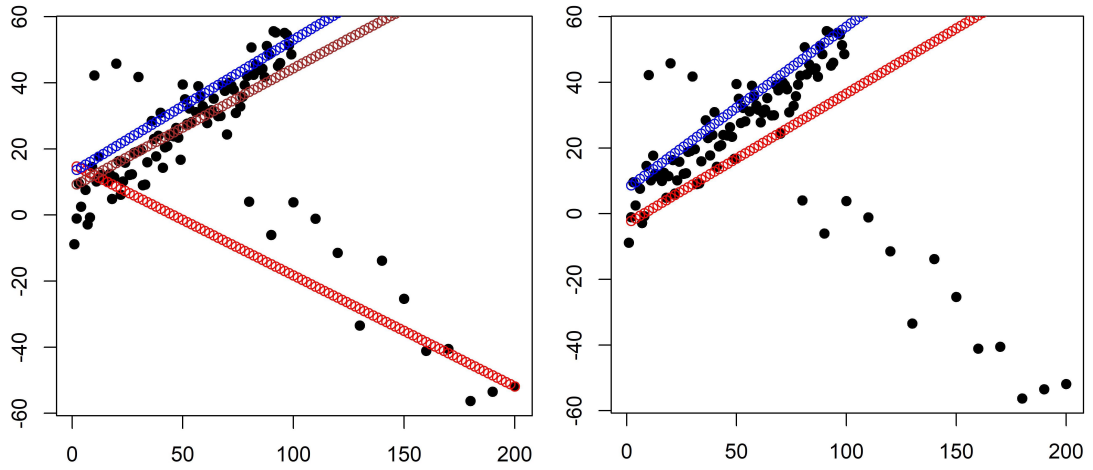


Figure 3 Simulation B of Section 4.3. Left: results of standard regression quantiles with $\tau = 0.15$ (bottom), $\tau = 0.5$ (middle) and $\tau = 0.85$ (top). Right: results of the QLWS estimator with $c = 0.005$ (top) and $c = 5$ (bottom).

LWS estimator has well known appealing properties (efficiency and at the same time high robustness) and there is a very good experience with its practical performance over real data. Numerical examples of Section 4 on three datasets reveal the potential of the newly proposed methods. So far, we have not accompanied the new proposals by theoretical results concerning the robustness or efficiency yet.

Based on the results of the examples, we can make the following conclusions for the QLWS estimator. It seems conceptually simple, comprehensible and able to attain very good approximation results. It is especially useful that it seems robust with respect to outliers as well as leverage points, where the latter (i.e. resistance to outlying values on the horizontal axis) is revealed in Figure 3. Apart from the newly proposed QLWS estimator, it is possible (and natural) to extend the definition to the IQLWS as a novel robust estimator of linear trend, similarly with the known TLS estimator. As the results in Table 1 reveal, the new interquantile estimator IQLWS yields results similar to the LWS. While the LWS is theoretically known to be efficient as well as robust, it means that the QLWS estimators (as the basis for computing the IQLWS) do not apparently lose efficiency or robustness.

Moreover, we intend to devote our future research to several open questions:

- More extensive computations of the novel methods, especially for more complex data (i.e. with a larger number of regressors),
- Theoretical investigation of properties of the novel methods,
- Analogous quantiles for multivariate data as an alternative to [17] (with possible applications to classification),
- Analogous quantiles in nonlinear regression (possibly in the context of multilayer perceptrons or radial basis function networks).

Appendix

Definition 2 (Ranks)**.** Let $X = (X_1, \ldots, X_n)^T$ represent a random vector with values in $(\mathbb{R}^n, \mathcal{B}^n)$, where $\mathcal{B}$ is the system of Borel sets on $\mathbb{R}$. Arranging X in ascending order, we obtain the vector of order statistics

$$X^n_{(1)} \leq X^n_{(2)} \leq \cdots \leq X^n_{(n)}, \tag{9}$$

where the upper index stresses that these are computed from n random variables. If no two observations are equal, the rank R_i of the i-th observation is defined by $X_i = X_{(R_i)}$ for $i = 1, \ldots, n$.

Definition 3 (Weight function)**.** Let $\psi : [0,1] \to [0,1]$ be a non-increasing and continuous function on $[0,1]$, let $\psi(0) = 1$ and $\psi(1) = 0$. Moreover, we assume that both one-sided derivatives of ψ exist in all points of $(0,1)$, that they are bounded by a common constant and we assume the existence of a finite left derivative in 0 and finite right derivative in point 1. Then the function ψ is called a weight function.

Acknowledgments

This work was supported by the Czech Science Foundation grants 19-05704S (J. Kalina) and 18-23827S (P. Vidnerová).

References

[1] Adrover, J., Maronna, R.A., & Yohai, V.J. (2004). Robust regression quantiles. *Journal of Statistical Planning and Inference*, 122, 187–202.

[2] Broniatowski, M., Jurečková, J., & Kalina, J. (2018). Likelihood ratio testing under measurement errors. *Entropy*, 20 (12), Article 966.

[3] Elamin, O.A. (2013). *Nonparametric kernel estimation methods for discrete conditional functions in econometrics*. Doctoral thesis. Manchester: The University of Manchester.

[4] Galarza Morales, C., Lachos Davila, V., Barbora Cabral, C., & Castro Cepero, L. (2017). Robust quantile regression using a generalized class of skewed distributions. *Stat*, 6, 113–130.

[5] Gutenbrunner, C. & Jurečková, J. (1992). Regression rank scores and regression quantiles. *Annals of Statistics*, 20, 305–330.

[6] Hájek, J. & Šidák, Z. (1967). *Theory of rank tests.* Prague: Academia.

[7] He, X. (1997). Quantile curves without crossing. *American Statistician*, 51, 186–192.

[8] Jurečková, J., Picek, J., & Schindler, M. (2019). *Robust statistical methods with R.* 2nd end. Boca Raton: CRC Press.

[9] Kalina, J. (2015). Three contributions to robust regression diagnostics. *Journal of Applied Mathematics, Statistics and Informatics*, 11, 69–78.

[10] Kalina, J. (2018). A robust pre-processing of BeadChip microarray images. *Biocybernetics and Biomedical Engineering*, 38, 556–563.

[11] Kalina, J. & Rensová, D. (2015). How to reduce dimensionality of data: Robustness point of view. *Serbian Journal of Management*, 10, 131–140.

[12] Koenker, R. (2005). *Quantile regression.* Cambridge: Cambridge University Press.

[13] Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, 9, 155–176.

[14] Koenker, R., Chernozhukov, V., He, X., & Peng, L. (2017). *Handbook of quantile regression.* Boca Raton: Chapman & Hall/CRC.

[15] Meier, V. (2011). *Econometric analysis of growth and convergence.* Doctoral thesis. Bielefeld: Universität Bielefeld.

[16] Procházka, B. (1988). Regression quantiles and trimmed least squares estimator in the nonlinear regression model. *Computational Statistics and Data Analysis*, 6, 385–391.

[17] Paindaveine, D. & Šiman, M. (2011). On directional multiple-output quantile regression. *Journal of Multivariate Analysis*, 102, 193–212.

[18] Rousseeuw, P.J. & van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery*, 12, 29–45.

[19] Víšek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47, 179–206.

Bankruptcy Prediction Based on Data Envelopment Analysis

Michaela Staňková[1], David Hampel[2]

Abstract. This article focuses on the classification based on data envelopment analysis for bankruptcy prediction of engineering companies in the EU. Data envelopment analysis is a technique developed in operational research that is typically used for efficiency evaluation. In this article, the data envelopment analysis method is used to find a "bankruptcy frontier", which makes it possible to classify the companies as active (healthy) or bankrupt. For this purpose an additive model is selected. We deal with 953 small and medium-sized engineering companies in the EU from which 51 companies bankrupted in 2014. All models are based on financial data of companies. With regards to financial theory, a set of 19 characteristics representing the four basic groups of financial indicators (i.e. solvency ratios, liquidity ratios, turnover ratios and profitability ratios) is chosen. The results show that the data envelopment analysis method could be a competitor in the area of bankruptcy prediction for other methods such as support vector machines or logistic regression.

Keywords: bankruptcy prediction, company accounting data, data envelopment analysis, linear programming.

JEL Classification: C44, C38
AMS Classification: 90B50, 90C08

1 Introduction

Evaluating the financial situation of company is a very important topic for decades in the academic or practical field of corporate finance. Knowledge of the company's financial situation is important not only for the top management, shareholders and creditors of the company in their decision-making, but also for other subjects. For example, because the financial distress of the company – especially for a larger number of bankrupt companies over the same period – brings serious problems such as typically unemployment. Unemployment rate is very important factor for the government. Therefore, there is a constant demand for ever more accurate and stable tool for predicting the company's financial situation.

Since the second half of the 20th century, the problematic of evaluation of the financial health of the company has begun to pay more attention. A myriad of models were developed to predict the financial situation of a company. All the models evaluating the financial situation of the company are based on the premise that companies have symptoms typical of bankruptcy for some time before this status. Based on the company's financial statements, it is possible to define many financial ratios that could predict the bankruptcy of the company. These financial indicators can be used to construct the model. The first well-known models such as the Altman Z-Score were based on multiple discriminant analysis, see [1]. The logistic regression models or probit models are also very popular in this area, see [5] or [20]. Less traditional methods, such as neural networks, decision trees, or support vector machines, are also used here, see [8], [13], [14] and [17]. In all these models, financial ratios play a significant role.

DEA is a technique developed in operational research. It is a special case of linear programming, which is currently used mainly for efficiency evaluation, like in [21]. The basis of the DEA method is stated in Farrell's work [7]. Later, different models were developed on this basis. The three most known DEA models are the CCR model by Charnes, Cooper, and Rhodes [11]; the BCC model by Banker, Charnes, and Cooper [4] and the Additive model (and later derived SBM models) by Charnes et al. [10]. DEA is non-parametric methods that compare the observed inputs and outputs of each unit with that of the most performing unit in the information dataset. The

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, michaela.stankova@mendelu.cz.
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, david.hampel.uso@mendelu.cz.

goal is to find an efficiency frontier, which is made up of efficient units. In the case of efficiency assessment, input variables are typical production inputs (such as labour and capital). Output variables include company outputs (i.e. production).

This article focuses on the process of identifying bankrupt companies using the data envelopment analysis method (DEA). Recent studies using the DEA method for this purpose are [12], [16], [18], [19] and [22]. The aim of this article is to evaluate the accuracy of models for bankruptcy prediction based on DEA method. In this case, the DEA methodology is used to create a "bankruptcy frontier", which makes it possible to classify the companies as active (healthy) or bankrupt. For this purpose we will use typical financial ratios as input and output variables. The results will be compared with other methods used to assess the financial health of the company.

2 Material and Methods

The annual accounting data are collected from the Orbis database. We deal with 953 engineering companies in the EU from which 51 companies bankrupted in 2014. To achieve more homogeneous data set, only small and medium-sized engineering companies are included. With regards to financial theory, a set of 19 characteristics representing the four basic groups of financial indicators (i.e. solvency ratios, liquidity ratios, turnover ratios and profitability ratios) are chosen. Using ratio indicators in the DEA can be problematic. In [9], is stated, that one of the main assumptions (related to production possibility set) in the definition of efficiency measure underlying DEA is the convexity axiom. It is problematic not to violate this assumption when ratios are used. However, ratios are the basis for assessing the financial health of a companies and they are therefore essential for these type of analyzes. Other assumptions related to production possibility set listed in [6] are fulfilled. In Tab. 1, there are medians of financial variables used in our analysis.

	Active companies			**Bankrupt companies**		
Variable median/Year	**2011**	**2012**	**2013**	**2011**	**2012**	**2013**
Current ratio	1.52	1.54	1.57	1.09	1.01	0.75
Quick ratio	0.98	1.01	1.05	0.67	0.59	0.41
Cash ratio	0.42	0.44	0.48	0.18	0.13	0.10
Cash flow liquidity	0.14	0.15	0.15	0.03	-0.03	-0.03
Net working capital (mln EUR)	15.03	15.76	16.19	0.15	0.01	-0.43
Working capital per employee (th EUR)	59.00	59.56	59.28	71.71	61.64	36.00
Return on assets (%)	3.95	4.18	3.82	-0.13	-3.64	-38.03
Profit margin (%)	4.66	4.39	4.62	0.82	-3.99	-65.77
EBIT Margin (%)	5.07	4.99	5.17	2.76	-1.73	-57.68
Net assets turnover	2.52	2.52	2.41	3.33	2.68	1.52
Stock turnover	5.61	5.75	5.86	3.16	2.75	2.57
Collection period (days)	66.00	63.89	63.66	111.73	142.57	151.82
Credit period (days)	46.86	42.93	43.03	97.94	114.01	230.94
Debt ratio (%)	66.88	64.9	63.64	93.14	97.45	138.58
Solvency ratio (%)	33.12	35.10	36.36	6.86	2.55	-18.42
Debt to equity ratio	1.96	1.76	1.69	9.14	7.69	-2.07
Interest cover	6.90	6.97	7.82	1.11	-0.78	-9.49
Debt repayment period	6.15	6.10	5.50	19.38	-2.26	-3.57
Fixed asset to equity capital ratio	1.28	1.32	1.30	0.32	0.14	-1.92

Table 1 Medians of used variables for active and bankrupt companies in particular periods

Only four financial ratios were selected as output variables. These variables are debt repayment period, debt ratio, credit period and debt to equity ratio. This combination of financial ratios was chosen because the output variables are variables that we want to maximize in the DEA model. Bankrupt companies may be expected to have a higher level of indebtedness (more precisely total debt, especially liabilities), so both debt ratio and debt to equity ratio will increase. The bankruptcy of the company will also affect with negative impact the operating cash flow, so the debt repayment period will also be prolonged. Similarly, companies will be more likely to

repay their own liabilities, so the credit period will be extended. The resting 15 financial indicators represent input variables. Here, on the contrary, it can be assumed that value of these variables in the case of bankrupt companies should be lower than in the case of active companies. It is a group of profitability ratios and liquidity ratios and the remaining indicators in the group of solvency ratios and turnover ratios.

From the three most known DEA models mentioned above, the additive model was selected (like in [12], [18] and [19]). Technical details about the additive model can be found in [6]. The advantage of this model is that it allows working with negative values in inputs and outputs, which is useful in bankruptcy assessment where financial ratios are often negative. In order to avoid the problem with the units of measurement, the input and output variables were standardized. Compared to classic CCR and BCC models, the additive model deals with the input and output slacks (i.e. input excesses and output shortfalls) directly in the objective function. In this type of DEA model, company's performance is determined by examining slacks only. This is another reason for choosing this model, according to [18]. Because an efficiency score measured by the ratio form depends upon measurement, the input-based efficiency score is different from that of the output-based one. In contrast, the additive model can avoid the problem associated with the ratio form.

Based on matrix of inputs (X) and matrix of outputs (Y) we can measure the performance of each company H by solving n times the following additive model:

$$\begin{aligned} &\max_{\lambda, s^-, s^+} z = e^T s^+ + e^T s^- \\ \text{subject to} \quad & X\lambda + s^- = x_H\,, \\ & Y\lambda - s^+ = y_H\,, \\ & e^T\lambda = 1\,, \\ & \lambda \geq 0,\ s^+ \geq 0,\ s^- \geq 0\,, \end{aligned} \tag{1}$$

where λ is the nonnegative weight vector, e^T is vector of ones, s^+ is a vector of output slacks, s^- is a vector of input slacks, x_H is a column vector of inputs of the company H, y_H is a column vector of outputs of the company H and z is a scalar.

In the case of an additive model adapted for bankruptcy prediction, company will appear on the "bankruptcy frontier" if the both slacks value s^+ and s^- are zero. This means that the frontier is created by the poor performers, ideally only the bankrupt companies. Otherwise (at least one slack positive), the company is not on the "bankruptcy frontier". The DEA models are constructed individually for the period of one, two and three years before the company's bankruptcy.

To evaluate the success of classification, we can calculate total accuracy (as a percentage of correctly classified subjects for all entities) and errors. Type I error evaluates the number of active companies which were classified as bankrupt companies to all active companies. Type II error shows how many bankrupt companies were falsely identified as active companies to all bankrupt companies. Due to the deterministic nature of the method, it is possible that some incorrectly classified bankrupt companies may be located near the frontier. These companies do not have the value of all slacks zero, but very close to zero. For this reason, it is advisable to construct the receiver operating characteristic (ROC). This curve can be calculated based on both error rates at different levels of slacks. Moreover, the value of area under this curve (AUC) also serves to evaluate the model's quality. More details on these calculations can be found in [13].

All the described procedures are performed in the MATLAB R2018b computing system with Statistics and Machine Learning Toolbox and Data Envelopment Analysis Toolbox (more details about this specific toolbox can be found in [2]) and DEA SolverPro (version 15).

3 Results

At first, we looked at the total number of companies located at the "bankruptcy frontier" and outside this frontier. The results of all three periods are summarized in Tab. 2.It has been found that with the approaching bankruptcy time, the number of companies forming the "bankruptcy frontier" is decreasing. In 2011, the total number of 121 companies appeared on the "bankruptcy frontier", but only 52 companies in 2013. The reduction in the number of companies on the frontier is therefore more than half. Also noteworthy is that in 2013 the ratio of the company appeared on the frontier and outside the frontier is almost the same as the ratio of bankrupt and active companies. Unfortunately, in all three monitored periods, there are some active companies that have appeared in the "bankruptcy frontier". Even in 2013, these 52 companies that appeared on the frontier are not only from a group of bankrupt companies. However, the numbers of active companies that appeared on the frontier have fallen

considerably over the years. In 2013, the ratio of active and bankrupt companies was one-to-one. Positive is the fact that over the years the number of bankrupt companies that appeared on the bankruptcy frontier is increasing. In 2011 (i.e. the period of three years before bankruptcy), nearly 41% of bankrupt companies were correctly classified. In 2012, this number rose to 49%. In the last of the monitored periods, it is more than 50% of the bankrupt companies.

Year	Companies	Bankrupt	Active	Total
2011	On frontier	21	100	121
	Not on frontier	30	802	832
2012	On frontier	25	45	70
	Not on frontier	26	857	883
2013	On frontier	26	26	52
	Not on frontier	25	876	901

Table 2 Summary of the DEA results

For evaluating the quality of these models, the values of total accuracy and also the values of both type I and II errors were calculated. Since the proportion of active and bankrupt companies is not balanced in the data set, it is advisable to highlight overall error rates for both active and bankrupt companies in order to prevent the loss of error margin classification of the less frequent companies that went bankrupt during observed period. That is why type I and II errors were calculated separately. The results of total accuracy and both type of errors can be found in Tab. 3.

Year	Accuracy	Type I error	Type II error
2011	0.8636	0.1109	0.5882
2012	0.9255	0.0499	0.5098
2013	0.9402	0.0288	0.4902

Table 3 Indicators of predictive quality

Total accuracy values for the period of the three years before bankruptcy are higher than 86%. This is a relatively high number. In 2013, over 94% of companies were correctly classified. Looking at the individual error rate, we can see that the increase in these accuracy values is due to the reduction of the error rate for active companies (i.e. type I error) mainly.

Fig. 1 represents the ROC curve for each of the three mentioned periods. With approaching bankruptcy the area under the ROC curve increases visibly. In the period of the three years before bankruptcy, the AUC value was 0.7743 (i.e. area under red curve). In period of the two years before bankruptcy (represented by the green ROC curve), the AUC increased to 0.8390. And in 2013, the AUC value is even 0.8811 (i.e. values related to the blue curve). Since AUC values can range between 0.5 and 1, the calculated AUC values are relatively high, indicating the good quality of these models.

4 Discussion

Based on the resulting accuracy and error rates, the DEA method appears to be an appropriate tool for assessing the financial health of an enterprise. In the studies [12] and [18] dealing with different industry (and based on another number and combination of variables), the results of the predictive ability were quite similar (total accuracy and both error rates are at a comparable level). The results of the predictive ability of the DEA models can be compared with those of other previous mentioned methods. For example, when compared to the study [17], it can be seen that the DEA model based on data coming from more than one years before the bankruptcy does not achieve qualities of support vector machines model (where total accuracy values equal 93%). However, all these DEA models have overcome their model based on multiple discriminant analysis, as well as models based on logistic regressions. In [17], the logit model achieved an accuracy of 85%. Model based on multiple discriminant analysis had reached the lowest accuracy, of less than 82%.

However, the most accurate comparison can be made with the study [20]. In this article, the same dataset was used to find the most suitable bankruptcy prediction model using method of logistic regression, support vector machines and decision trees. In that paper, a large number of models had been estimated – some with better

results in total accuracy and type I and II error (and associated AUC values) than our DEA models, but some were significantly worse. The best of models estimated in [20] achieved a total accuracy of 95% in 2011, 96% in 2012 and almost 99% in 2013. Similarly, these models achieved higher AUC values; more precisely 0.91 in 2011, 0.93 in 2012 and 0.98 in 2013. The accuracy of DEA models could be compared with other methods like multiple-criteria decision-making, neural networks or *k*-nearest neighbor. These are areas of upcoming research.

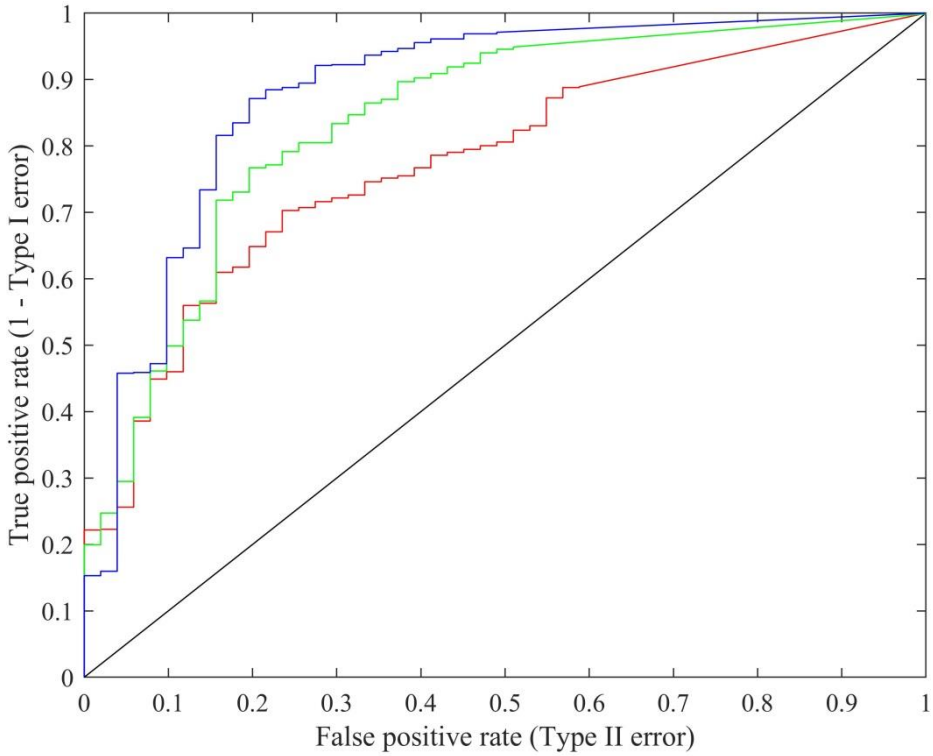


Figure 1 ROC curve in 2011 (red curve), 2012 (green curve) and 2013 (blue curve)

Although the DEA model does not achieve the same quality as the best of the models in [18], there is still the possibility of examining the results at various settings. It is possible to focus in more detail on other types of non-radial models, such as SBM models, or super-efficiency models. The super-efficiency models themselves can be processed in both radial and non-radial measuring modes (i.e. both additive and SBM variants). In such models, it would be advisable to verify the threshold that is used to distinguish between active and bankruptcy businesses. Another option is to use so-called bounded DEA models defined by [3], and combine the traditional optimistic and pessimistic point of view of production possibility set. Direct comparison with this approach is a goal for further research.

The resulting accuracy (and AUC values) of the DEA models could be even higher if the bankrupt companies were more frequent in the data set. Conversely, thanks to this, a data set with a predominance of active companies can be described as more realistic because it better reflects the situation on today's market which is populated far more densely by active companies than by those that are on the brink of bankruptcy.

5 Conclusion

The results clearly show that with increasing distance from bankruptcy the predictive ability of examined models decreases. The resulting total accuracy and AUC values are higher for data coming from one year before bankruptcy. Conversely, both errors are reduced over time. The DEA method is not primarily designed for the purposes of bankruptcy prediction. But compared to other methods it has promising results for assessing the company's financial health. The analysis could be extended in a variety of ways. In addition to looking at the predictive abilities of other types of non-radial DEA models, it would be possible to look for a more appropriate cutoff point for company classification.

Acknowledgements

This article was supported by the grant No. PEF/DP/2019035 of the Internal Grant Agency PEF MENDELU.

References

[1] Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23, 589–609.
[2] Álvarez, I. C., Barbero, J., Zofío, J. L. (2016). *A Data Envelopment Analysis Toolbox for MATLAB, Working Papers in Economic Theory 2016/03*, Universidad Autónoma de Madrid (Spain): Department of Economic Analysis.
[3] Azizi, H., Ajirlu, S. F. (2010). Measurement of overall performances of decision-making units using ideal and anti-ideal decision-making units. *Computers & Industrial Engineering*, 59(3), 411-418.
[4] Banker, R., Charnes, A., Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078–1092.
[5] Boratyńska, K., Grzegorzewska, E. (2018). Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches. *Journal of Business Research*, 89, 175–181.
[6] Cooper, W. W., Seiford, L. M., Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. 2nd edition. New York: Springer Science & Business Media.
[7] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3), 253–290.
[8] Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Application*, 117, 287–299.
[9] Emrouznejad A., Amin, G. R. (2009). DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling*, 33(1), 486–498.
[10] Charnes, A., Cooper, W. W., Golany, B., Seiford, L., Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1–2), 91–107.
[11] Charnes, A., Cooper, W., Rhodes, E. (1987). Measuring the efficiency of decision making units. *Management science*, 2(6), 429–444.
[12] Janová, J., Vavřina, J., Hampel, D. (2012). DEA as a tool for bankruptcy assessment: the agribusiness case study. *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012. Karviná: Silesian University in Opava, School of Business Administration in Karviná*, 379–383.
[13] Klepáč, V. Hampel, D. (2018). Predicting Bankruptcy of Manufacturing Companies in EU. *Economics and management*, 21(1), 159–174.
[14] Klepáč, V. Hampel, D. (2017). Predicting financial distress of agriculture companies in EU. *Agricultural Economics*, 63, 347–355.
[15] Klepáč, V. Hampel, D. (2016). Prediction of Bankruptcy with SVM Classifiers Among Retail Business Companies in EU. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 64, 627–634.
[16] Mousavi, M. M., Ouenniche, J., Xu, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, 42, 64–75.
[17] Niknya, A., Darabi, R., Fard, V. H. R. (2013). Financial Distress Prediction of Tehran Stock Exchange Companies Using Support Vector Machines. *European Online Journal of Natural and Social Science*, 2(3), 2878–2893.
[18] Premachandra, I. M., Bhabra, G. S., Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, 193(2), 412–424.
[19] Premachadra, I. M., Chen, Y., Watson, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega*, 39(6), 620–626.
[20] Staňková, M., Hampel, D. (2018). Bankruptcy prediction of engineering companies in the EU using classification methods. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 66(5), 1347–1356.
[21] Staňková, M., Hampel, D. (2018). Efficiency Comparison in the Development of Building Projects Sector. *Mathematical Methods in Economics 2018: Conference Proceedings. MatfyzPress: Praha*, 503-508.
[22] Zhiyong, L., Crook, J., Andreeva G. (2017). Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications*, 80, 94–106.

Inter-Branch Comparison of Cluster Company Performance Using Malmquist Index

Miroslav Žižka[1]

Abstract. The article evaluates the differences between the financial performance of clustered companies in two sectors for the period of 2009-2016. The first sample consists of 16 member companies that form the core CLUTEX cluster of technical textiles. The second sample includes 22 subjects representing NANOPROGRESS - cluster core of nanotechnology. Financial performance development was analyzed using the Adjacent Malmquist Index based on window scores. Window score values were obtained using Data Envelopment Analysis. Staff number, total assets and a long-term capital were used as model inputs. The economic value added was the model output. For each sample, the Malmquist Index and its components were determined - technical efficiency change and technological change. For scale efficiency, models with both constant and variable returns to scale were applied. The results of the analysis show that the average annual growth in company performance in the nanotechnology cluster is significantly higher than in the case of the textile cluster. However, the fact that the growth of performance is driven by technological change, that is to say, innovation, is a positive factor for the existence of a textile cluster. In the nanotechnology cluster, the main source of performance growth is the improvement of internal technical efficiency.

Keywords: Adjacent Malmquist Index, economic value added, financial performance, cluster of technical textiles, cluster of nanotechnology.

JEL Classification: C61, L25, L67
AMS Classification: 90B90, 90C90

1 Introduction

Clusters are considered to be one of the drivers of regional competitiveness. The concept of "cluster" or "industry cluster" is related mainly to the work of Michael Porter. According to Porter, the cluster can be understood as *„geographic concentrations of interconnected companies and institutions in a particular field. Clusters encompass an array of linked industries and other entities important to competition. They include, for example, suppliers of specialized inputs such as components, machinery, and services, and providers of specialized infrastructure. Clusters also often extend downstream to channels and customers and laterally to manufacturers of complementary products and to companies in industries related by skills, technologies, or common inputs. Finally, many clusters include governmental and other institutions such as universities, standards-setting agencies, think tanks, vocational training providers, and trade associations, that provide specialized training, education, information, research, and technical support"* [15]. Clusters can arise on the basis of natural market developments or as a result of a cluster initiative by some, usually governmental institutions. In the Czech Republic, clusters have been supported since 2004 through several operational programs with funding from the EU Structural Funds. The main purpose of supporting cluster initiatives was to strengthen cooperation between the business community (especially SMEs) and the research community in the field of joint research, development and innovation activities. As a result, knowledge-based economics should develop. An important aspect is that collective research must represent an economic benefit for small and medium-sized enterprises [13]. At present (February 2019) there are 85 cluster organizations in the Czech Republic, according to a survey conducted at the Technical University of Liberec. The aim of the research is to find out whether the cluster's establishment brings economic benefit to the participating companies in the longer term.

The research underlying this article was conducted in two cluster organizations, the member of one of them is also the Technical University of Liberec. CLUTEX, the cluster of technical fabrics represents the traditional industrial sector. It was founded in 2006 and operates in three regions - Liberec, Hradec Králové and Pardubice. The aim of the CLUTEX cluster is to create optimal conditions for the transfer of technology, to ensure the innovation of higher orders and the development of entrepreneurship in the field of research, development and production of technical textiles, including materials and semi-finished products used for their production [4].

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec 1, miroslav.zizka@tul.cz.

The NANOPROGRESS cluster originated in 2010, operates in the same region as the previous cluster, but represents the modern nanotechnology industry. The NANOPROGRESS cluster focuses on researching and developing functional nanofiber structures and their industrial and medical applications [12].

The aim of the article is to find out the differences in the financial performance of companies in both clusters in relation to the industry in the time series of 2009-2016 and to identify the causes that contributed to the change in performance. Financial value added was used as a measure of financial performance. The article builds on previous research [16], however, in the new one, a research file has been added and the time series extended by the year 2016.

2 Malmquist Index

The Malmquist Index (furthermore MI) is one of the productivity indexes. It measures a total factor productivity change or company or branch performance between individual periods. Several indexes can be used to quantify these changes. It is described, for example, by Coelli [5]. If the effect of changing input and output prices is not considered, then the main source of change in productivity is technological change and change in efficiency (technical, scale) over the period. MI uses input and output distance functions. Various econometric methods or mathematical programming methods can be used to determine distance functions. Data Envelopment Analysis is used in this article. MI is designed to measure the radial distance of the observed input and output vectors in two periods, most often (t) and (t + 1), with a given reference technology. However, the period distance may be wider, then it is a window-analysis. Models can be oriented to output or input. The MI results differ according to the orientation used.

This article uses the so-called Adjacent Malmquist Index based on DEA window scores F, which represent a special type of distance functions. The MI calculation based on DEA window-analysis is based on the principle that each unit (a company) in a given period is considered independent. Thus, the performance of one and the same unit between the different periods a and b is compared. Such an approach is particularly useful in the case of a small number of units where the amount of data increases.

The width of the window then determines the length of the analyzed period. If the width of the window is w = 1, then it is not necessary to indicate it in the formula, since the two adjacent periods are compared. It is assumed that there are no technological changes inside the window [14]. In the present study, the adjacent Malmquist index with window width w = 1 was used. The computational relationships are given in the formulas (2) to (4) and taken from the article [1]. As mentioned above, MI breaks down the change in productivity into a change in efficiency and technological change between two periods, see the relationship (1). Technical efficiency change *EFFCH* reflects the extent to which the unit approaches the most efficient production frontier. This change reflects the internal efforts of the unit to improve performance [9]. Technological change *TECH* is characterized by an efficiency frontier shift. Technological change occurs mainly due to innovation in the industrial sector. The computational relations for determining the individual MI components are given in equations (3) and (4).

In general, it is desirable for index values to be higher than one. In this case, there is an increase in productivity, efficiency and innovation. Changing *EFFCH* technical efficiency can be further decomposed into the product of *PECH* change of pure technical efficiency (under variable returns to scale conditions) and *SECH* scale efficiency change. Technological change *TECH* can be expressed as the product of three components *OBITECH, IBITECH* and *TECHM*. The *OBITECH* and *IBITECH* components evaluate the change in the input/output ratio in relation to the efficient frontier between the periods. Technical change magnitude *TECHM* measures the relative distance of efficient frontiers between two periods under the input and output neutrality conditions.

$$MI_q(a,b) = EFFCH_q TECH_q = (PECH_q SECH_q)(TECHM_q OBITECH_q IBITECH_q) \quad (1)$$

$$MI_q(a,b) = \sqrt{\frac{F^q_{a_w,b} F^q_{b_w,b}}{F^q_{a_w,a} F^q_{b_w,a}}} \quad (2)$$

$$EFFCH_q(a,b) = \frac{F^q_{b_w,b}}{F^q_{a_w,a}} \quad (3)$$

$$TECH_q(a,b) = \sqrt{\frac{F^q_{a_w,b} F^q_{a_w,a}}{F^q_{b_w,b} F^q_{b_w,a}}} \quad (4)$$

$$F^q_{a_w,b} = \min_{\beta,\lambda} \beta \quad (5)$$

Since the output used in the model - economic value added - can have both positive and negative values, a modified radial measure (VRM) model [3] was used to calculate the window scores. Equations (6) of the input-

oriented VRM model under CRS conditions are listed below. The modified model uses absolute values of inputs and outputs instead of their actual values. The difference $(1 - \beta)$ expresses the efficiency measured by the DMU. The quantity β is then a measure of inefficiency. It indicates the degree of improvement needed to reach the frontier when applying a proportional reduction of inputs [3]. In relations (5), **X** is the input matrix, **Y** is the matrix of outputs, $\boldsymbol{x_q}$ is the input vector of the unit q, $\boldsymbol{y_q}$ is the output vector of the unit q and λ is the vector of weights assigned to each unit.

$$\begin{gathered} \max \beta \\ \boldsymbol{X}\lambda + \beta|\boldsymbol{x}_q| \leq \boldsymbol{x}_q \\ \boldsymbol{Y}\lambda \geq \boldsymbol{y}_q \\ \lambda \geq 0 \end{gathered} \tag{5}$$

3 Data and methodology

For the research, data from the balance sheets, the profit and loss statements and the supplements to the financial statements of the member companies of CLUTEX and NANOPROGRESS clusters for the period of 2009-2016 were used. The accounting data were obtained from the MagnusWeb database [2]. In three cases, the accounting data from some years were missing from the database, and the financial statements had to be traced back in the collection of commercial register documents and rewritten into tables in MS Excel. Since the MagnusWeb database from 2018 no longer contains historical data on the number of company employees, this data had to be searched for in the collection of documents of the register of companies. Supplements to the financial statements, or annual company reports for 2009-2016 also served as a source of data. The beginning of the time series was chosen with respect to the year of the creation of both clusters. The end of the time series is affected by the availability of financial statements. The financial statements for 2017 for many companies are not yet published.

The research was carried out in the following steps:

1. **Creation of a list of CLUTEX and NANOPROGRESS cluster cores** – CLUTEX cluster has currently 35 members (February 2019). Compared to 2018, the number of members increased by five. The business activities of individual member entities are quite different. From the point of view of Data Envelopment Analysis, it is essential that the units can be considered homogeneous. The core of the cluster, which is made up of companies in NACE 13200, 13900 and 14100 meets this assumption. There are 19 core companies, out of which 16 companies have a complete time series of financial statements and represent the first research sample. The second research sample is made up of member companies of the NANOPROGRESS cluster. This cluster currently has 38 members from industry, health, education, consultancy and trade. The core activities are the NACE 720 and 721 branches. This activity is performed by 23 cluster members. A complete time series of financial statements was available for 22 entities that form the second research sample.

2. **Determining inputs and outputs for Data Envelopment Analysis** - inputs of the model are the number of employees, total assets and long-term capital of the enterprise (the sum of equity, long-term bonds issued and long-term payables to credit institutions). The output is the economic value added (EVA), which represents net operating profit after deducting capital costs. EVA was calculated according to the methodology of the Ministry of Industry and Trade [10] as the product of equity and the spread between return on equity and alternative cost of equity. EVA can have both positive and negative values. To create value for owners, it is desirable for EVA to get positive numbers. To eliminate the impact of price changes, financial data was used at constant prices in 2009.

   For the first cluster, the industrial producer price index in the textile, clothing and leather industries was used to convert current prices. For the second cluster, the price index of market services in other professional, scientific and technical services was used. Both indexes are published by the CZSO in its public database, see Table 1 [6].

Price index	2010	2011	2012	2013	2014	2015	2016
Textile, clothing and leather industry	1.005	1.053	1.009	1.000	1.024	1.001	1.001
Other professional, scientific and technical services	1.001	1.003	1.002	0.999	0.992	1.002	1.004

Table 1 Prices of Industry and Market Services [6]

3. **Determining the Malmquist Index and its components** – for each member enterprise (DMU) and period, the adjacent Malmquist Index values based on DEA window scores were determined, see previous chapter. The model is input-oriented, with radial distances, both constant CRS and returns to scale (VRS) were used to calculate the scale efficiency. For each period and each research sample (CLUTEX, NANOPROGRESS) the

geometric means of the Malmquist index and its components were then determined. The MaxDEA 7 Ultra software was used for the calculations.

4. **Comparison of the Malmquist Index and its components across branches** - since the values of the individual indicators do not have a normal distribution, which was verified using the Shapiro-Wilk test on the 5% significance level, it was necessary to compare the differences in the financial performance change between CLUTEX clusters companies and in the NANOPROGRESS cluster using a non-parametric Mann-Whitney (Wilcoxon) W-test. The test statistic gives the maximum distance between the cumulative distributions of the two samples.

4 Research results

The MI values and the individual components for both clusters are shown in Tables 2 and 3. The tables show that the overall change in the financial performance of member companies in both clusters for 2009-2016 is significantly different. In the textile sector, performance increased on average by only around 2% per year, by contrast, in the nanotechnology sector it increased on average by 55% per year. The difference is significant on the level of 5% (see Table 4). The reasons for growth are also different. In CLUTEX cluster, the growth in cluster company performance was driven solely by technological advances, averaging roughly 3% per year. The technological change in both sectors was caused in particular by changing the proportion of outputs and inputs (OBITECH component). There was no improvement in the technical efficiency of member companies in CLUTEX cluster over the entire period, there was even observed a small decrease of about 1.5% per year. In contrast, in the nanotechnology sector, performance growth was driven mainly by improving the internal efficiency of member companies. On average, internal efficiency in this sector grew by 51% per year. In the nanotechnology sector, the combined effect of pure technical efficiency and scale efficiency contributed to the improvement of internal efficiency. Conversely, in the textile sector, scale efficiency on average decreased significantly.

Year	MI_CRS	EFFCH_CRS	PECH_VRS	SECH	TECH	OBITECH	IBITECH	TECHM
2010/2009	0.9425	0.8883	2.3509	0.3778	1.0610	1.0610	1.0000	1.0000
2011/2010	1.1867	0.7606	2.5793	0.2949	1.5602	1.0567	1.0052	1.4689
2012/2011	0.9552	1.1947	0.9563	1.2493	0.7995	1.0247	1.0233	0.7625
2013/2012	1.0534	1.1463	2.0877	0.5491	0.9189	1.0106	1.0111	0.8994
2014/2013	0.9210	0.3876	0.7039	0.5507	2.3759	1.2880	1.0009	1.8430
2015/2014	0.8734	2.4840	2.0193	1.2301	0.3516	1.0086	1.0331	0.3374
2016/2015	1.2570	1.0229	1.9565	0.5228	1.2289	0.9482	1.0400	1.2462
GEOMEAN	**1.0186**	**0.9868**	**1.6526**	**0.5971**	**1.0322**	**1.0524**	**1.0161**	**0.9652**
MEDIAN	**1.0000**	**1.0000**	**1.1818**	**0.8166**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Table 2 Malmquist Index summary of annual means in the CLUTEX cluster

Year	MI_CRS	EFFCH_CRS	PECH_VRS	SECH	TECH	OBITECH	IBITECH	TECHM
2010/2009	1.2518	1.3781	1.1312	1.2182	0.9083	1.0476	1.0236	0.8471
2011/2010	2.0980	1.2852	0.9490	1.3543	1.6324	1.3131	0.9996	1.2437
2012/2011	1.1975	1.4029	1.2718	1.1031	0.8536	1.0272	1.0380	0.8006
2013/2012	4.2482	4.1122	1.7388	2.3650	1.0331	1.0353	1.0208	0.9776
2014/2013	1.5227	1.5833	1.4553	1.0880	0.9617	1.0162	1.0196	0.9282
2015/2014	0.8048	0.9546	1.0684	0.8935	0.8431	1.0625	1.0147	0.7820
2016/2015	1.2942	1.1180	1.0635	1.0512	1.1576	1.0668	1.0107	1.0737
GEOMEAN	**1.5486**	**1.5051**	**1.2168**	**1.2370**	**1.0289**	**1.0775**	**1.0181**	**0.9379**
MEDIAN	**1.1892**	**1.1726**	**1.0529**	**1.0534**	**1**	**1**	**1**	**1**

Table 3 Malmquist Index summary of annual means in the NANOPROGRESS cluster

From the point of view of development in individual periods, we can observe the fluctuating amount of the MI in both branches. However, in the sample of member companies of the NANOPROGRESS cluster the fluctuation was more pronounced, see Figure 1. It can be assumed that the performance of the business sphere is affected by the situation in the national economy. The Czech economy underwent recession in 2012-2013. Since 2014, the Czech economy has been continuously growing. In the sample of textile companies, there is an obvious decrease in financial performance between the years 2011 and 2012 of about 4.5%. At the same time, however, the internal efficiency of textile enterprises was increasing during this period. This means that in times of crisis companies were searching for internal savings in the organization of their business. Technological efficiency, on the other hand, was decreasing, which can be explained by the fact that companies limited their innovation. After a slight improvement in performance in the 2012/13 period, the downward trend continued until 2015. The financial performance of the textile companies that is predominantly driven by technological change was recovering only in

the last period under observation. In the second sample of nanotechnology companies, year-on-year fluctuations in MI are more pronounced, but even in times of economic recession, financial performance was growing. For both research samples there is a rather surprising year-on-year decline in performance in the period of 2015/2014. In the textile industry, the worsening economic results of one of the largest companies - VEBA, which resulted in the opening of its insolvency proceedings, can be identified as a cause. However, in the sample of nanotechnology companies, there was a year-on-year deterioration in the economic outcomes of several companies with no clear common cause. It can be a random swing. Interestingly, even in the nanotechnology branch, in the period of the economic crisis, there was a search for internal sources of savings and the reduction of technological efficiency, that is to say, reduction of innovations.

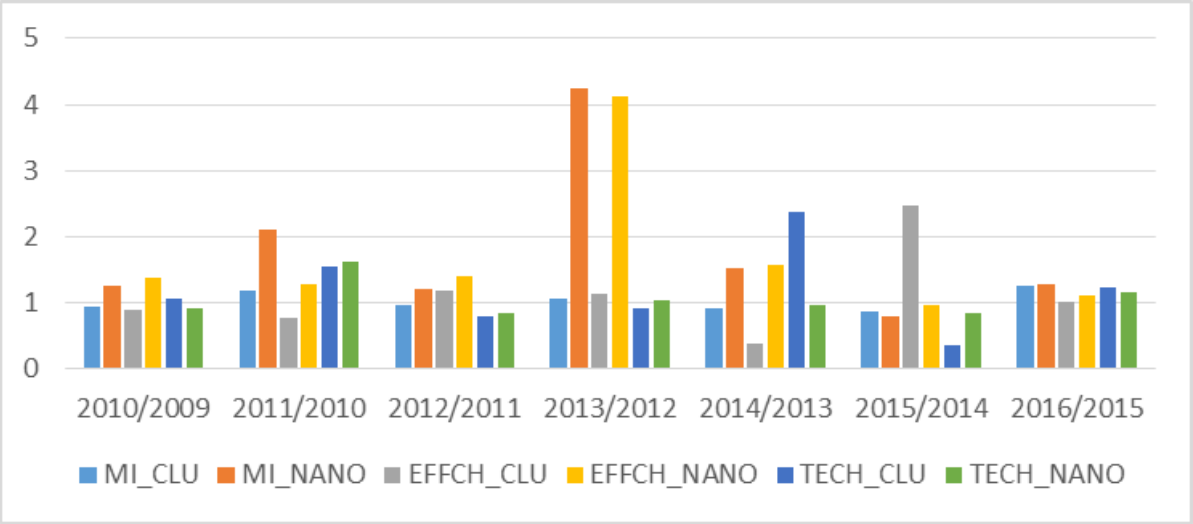


Figure 1 Comparison of Malmquist Indices and its components among clusters

Table 4 gives information about the significance of differences in size of MI and its components between the two research files. For the 2009-2016 period, it can be concluded that there are significant differences in the change in financial performance and its components between CLUTEX and NANOPROGRESS clusters.

Year	MI_CRS	EFFCH_CRS	PECH_VRS	SECH	TECH	TECHM
2010/2009	233.0	236.0	148.5	264.5	127.0	144.0
	0.0421	**0.0334**	0.5594	**0.0021**	0.0693	0.1281
2011/2010	264.0	267.0	99.0	326.0	171.0	149.0
	0.0028	**0.0019**	**0.0344**	**<0.0001**	0.9266	0.4067
2012/2011	217.5	217.5	133.0	229.0	174.5	175.5
	0.1264	0.1264	0.2896	0.0633	0.8200	0.7906
2013/2012	285.5	281.5	139.0	287.0	213.0	186.0
	0.0003	**0.0004**	0.3817	**0.0003**	0.0811	0.4563
2014/2013	266.0	284.0	228.5	285.0	95.0	117.0
	0.0025	**0.0003**	0.0657	**0.0004**	**0.0075**	**0.0159**
2015/2014	109.0	67.0	43.0	176.0	214.5	193.0
	0.0709	**0.0018**	**0.0001**	0.8179	0.0982	0.3369
2016/2015	154.0	167.0	90.5	264.5	150.5	129.0
	0.6749	0.9875	**0.0176**	**0.0032**	0.5190	0.1013
2009-2016	10841.0	10919.0	6861.5	13303.0	8091.0	7799.5
	0.0001	**0.0001**	**0.0077**	**<0.0001**	0.4200	0.1091

Table 4 Mann-Whitney (Wilcoxon) W-test, W statistic (P-Values) – CLUTEX vs. NANOPROGRESS

5 Conclusion

Based on the research conducted, it can be concluded that the financial performance of the member companies in the CLUTEX and NANOPROGRESS clusters developed differently. The average growth in financial performance in the textile cluster was significantly slower than in the nanotechnology cluster. Also, the economic crisis in 2012-13 hit the textile industry more strongly than the second branch of advanced technologies. It can be positively assessed that the growth of financial performance in the textile cluster is driven by innovation, that is, technological

change. This can be attributed to the spillover of knowledge between companies and research organizations gathered in the CLUTEX cluster. Scale efficiency in the textile sector is decreasing. This means that textile companies do not work at an optimum size. The textile industry has recorded a significant decline since 1990 in the Czech Republic in terms of both employment and capacities. Between 1991 and 2016, the number of employees in the textile industry decreased from 161 thousand to only 52 thousand [7, 8]. Many textile companies are still struggling with insufficiently utilized production capacities.

The results in financial performance of nanotechnology companies clearly show that it is a dynamically developing sector with a high average annual growth rate. Performance growth is also driven by improving internal technical efficiency, which means by organizational and managerial changes. Technological change expressing the shift of the efficient frontier, or technological progress, is not more pronounced than in the previous textile branch. The scale efficiency value indicates that even companies in nano-industrial branch are not working well in the optimum range. These are mostly young businesses that are in the conditions of increasing returns to scale. This is probably also the reason why there is still potential for improving internal organization with a positive impact on financial performance.

Further research in the area of financial performance monitoring will focus, in addition to obtaining numerical values for the next period, on the mapping of companies in other branches where cluster organizations exist. The aim of this research will be to reveal the impact of different industrial sectors on the development of financial performance and its components.

Acknowledgements

Supported by the grant No. GA18-01144S „An empirical study of the existence of clusters and their effect on the performance of member enterprises“ of the Czech Science Foundation.

References

[1] Asmild, M., Paradi, J.C., Aggarwall, V., and Schaffnit, C.: Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry. *Journal of Productivity Analysis* **21** (2004), 67–89.
[2] Bisnode: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2019 [cit. 2019-01-30]. Available at: https://magnusweb.bisnode.cz.
[3] Cheng, G., Zervopoulos, P., and Qian, Z.: A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research* **225** (2013) 100–105.
[4] CLUTEX: Stanovy klastru CLUTEX [online]. In: *Sbírka listin veřejného rejstříku.* Ministry of Justice CR, Prague, 2014 [cit. 2019-02-14]. Available at: https://or.justice.cz/ias/ui/rejstrik
[5] Coelli, T.J., Prasada Rao, D.S., O'Donnell, Ch.J., and Battese, G.E.: *An Introduction to Efficiency and Productivity Analysis.* Springer Science Business Media, New York, 2005.
[6] CZSO: *Public Database – Producer Prices* [online]. Czech Statistical Office, Prague, 2019 [cit. 2019-02-15]. Available at: https://vdb.czso.cz.
[7] CZSO: *Statistical Yerbook of the Czech Republic 1993*. Czech Statistical Office, Prague, 1993.
[8] CZSO: *Statistical Yearbook of the Czech Republic – 2017* [online]. Czech Statistical Office, Prague, 2017 [cit. 2018-03-17]. Available at: https://www.czso.cz/csu/czso/statistical-yearbook-of-the-czech-republic.
[9] Li, Z., Crook, J., and Andreeva, G.: Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems With Applications* **80** (2017), 94–106.
[10] MPO: *Finanční analýza podnikové sféry za rok 2017.* Ministry of Industry and Trade, Prague, 2018.
[11] MSP ČR: *Veřejný restřík a Sbírka listin* [online]. Ministry of Justice CR, Prague, 2015 [cit. 2019-01-20]. Available at: https://or.justice.cz/ias/ui/rejstrik
[12] Nanoprogress: *The Nanotechnology Cluster* [online]. Nanoprogress, Pardubice, 2017 [cit. 2019-02-14]. Available at: http://nanoprogress.eu/en/home-page-5/cluster-2/about-us
[13] OPPIK: *Operační program Podnikání a inovace pro konkurenceschopnost – Program Spolupráce* [online]. Enovation, Prague, 2018 [cit. 2019-02-14]. Available at: https://www.oppik.cz/dotacni-programy/spoluprace
[14] Palečková, I.: Application of Window Malmquist Index for Examination of Efficiency Change of Czech Commercial Banks. *DANUBE: Law and Economics Review* **8** (2017), 173–190.
[15] Porter, M. E.: Clusters and the new economics of competition. *Harward Business Review* **76** (1998), 77–90.
[16] Žižka, M.: Assessment of Cluster Benefits to Increase Financial Performance of Companies: Malmquist Index Approach.. In: *36th International Conference Mathematical Methods In Economics, Conference Proceedings* (Váchová, L., Kratochvíl, V., eds.). MatfyzPress, Prague, 2018, 648–653.

Relationship between Output and Unemployment

Andrea Čížků[1]

Abstract. The paper models relation between output and unemployment known as Okun's law for Germany (1991Q1-2017Q4) and Spain (1995Q1-2017Q4) by applying unobserved components methodology. The model is estimated by maximizing the likelihood function which is constructed by applying Kalman filter algorithm. Cyclical unemployment rate turned out to be driven to a great extent by cyclical properties of output. Estimated Okun's law coefficients for Germany are in line with other empirical studies. Nonetheless, much higher influence of output gap on unemployment rate gap is detected for Spain indicating substantial heterogeneity regarding the strength of the Okun's law among individual European countries.

Keywords: Okun's law, unobserved components, Kalman filter, maximum likelihood.

JEL Classification: C32, E32
AMS Classification: 91G70

1 Introduction

Correlation between output and unemployment is widely accepted empirical regularity due to Okun's [18] seminal contribution. Okun's law is macroeconomic empirical rule already well established in empirical economic literature. It postulates negative correlation between cyclical unemployment rate and cyclical component of output on a macroeconomic level. The question arises as to how to extract cyclical components of these two economic variables. Hodrick, Prescott [14] (HP) filter or Harvey's [12] unobserved components (UC) approach are the most commonly applied methodologies. Harvey, Jaeger [13] showed that UC methodology of detrending economic time series is superior to mechanical statistical tools like HP filter.

Firstly, the main goal of the paper is to describe the relation between output gap and cyclical unemployment rate by applying Clark's [9] bivariate UC model. This task is performed not for aggregated European data as by Berger [3], Berger, Everaert [4] or Orlandi, Pichelmann [19], but for two fundamentally different economies - Germany and Spain. Whereas our estimates of Okun's coefficients for Germany are in line with other empirical studies, this is not the case for Spain indicating considerable uncertainty about the strength of this relation even in this unstable economy. Secondly, structural (in)stability of the Okun's law is also discussed, which is a relevant question especially during the current economic crisis that begun in 2008.

The structure of the paper is as follows. The model is formulated in chapter 2. Data is described in chapter 3. Econometric methodology is summarized in chapter 4 and results from econometric estimation are discussed in section 5. The final chapter 6 summarizes main findings and concludes.

2 Model

Clark's [9] bivariate unobserved components model decomposing real GDP and unemployment rate into trend and a cycle component is described in this chapter. Output is decomposed as follows:

$$y_t = n_t + x_t\,, \tag{1}$$

$$n_t = g_{t-1} + n_{t-1} + v_t\,,\ v_t \sim i.i.d.N\left(0, \sigma_v^2\right), \tag{2}$$

$$g_t = g_{t-1} + w_t\,,\ w_t \sim i.i.d.N\left(0, \sigma_w^2\right), \tag{3}$$

$$x_t = \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + e_t\,,\ e_t \sim i.i.d.N\left(0, \sigma_e^2\right), \tag{4}$$

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67, Prague 3, Czech Republic, andrea.cizku@vse.cz.

where y_t is the logarithm of real GDP,

n_t describe a stochastic trend component,

g_t is the growth rate of the trend component,

x_t represents a stationary cyclical component,

v_t, w_t, e_t are independent white noise processes.

The autoregressive process of order two was chosen in the equation (4). This is a commonly applied assumption in empirical literature. It is a parsimonious way to model cyclical dynamics as it allows the output gap to exhibit cyclical movements.

Clark [9] extended this univariate model for output decomposition into a bivariate model of real GDP and unemployment. Unemployment rate decomposition is given as follows:

$$U_t = L_t + C_t, \tag{5}$$

$$L_t = L_{t-1} + \varepsilon_t,\ \varepsilon_t \sim i.i.d.N\left(0, \sigma_\varepsilon^2\right), \tag{6}$$

$$C_t = \alpha_0 \cdot x_t + \alpha_1 \cdot x_{t-1} + \alpha_2 \cdot x_{t-2} + \eta_t,\ \eta_t \sim i.i.d.N\left(0, \sigma_\eta^2\right), \tag{7}$$

where L_t is a trend component of unemployment rate,

C_t is a stationary component of unemployment rate,

ε_t, η_t are independent white noise processes.

The relation (7) is interpreted as Okun's law postulating negative relation between output gap and unemployment rate gap. This equation is a slight modification of the Okun's law formulated by Clark [9] as he assumed that only current output gap x_t and output gap lagged one period x_{t-1} have influence on cyclical unemployment rate C_t. Lagged values of output gap in the Okun's law (7) reflect that unemployment rate is a lagging indicator of business cycle. Including output gap lagged two periods x_{t-2} in the Okun's law was suggested by Kim, Nelson [15].

Note that output and unemployment rate are modeled simultaneously, not separately. Model of output decomposition (1)-(4) is interconnected with unemployment rate decomposition (5)-(7) by Okun's law (7) which postulates that output gap influence cyclical unemployment rate. The assumed causality goes from output to unemployment which is a lagging indicator. Econometric estimation of unobserved cyclical components of output and unemployment rate therefore utilizes this structural relationship between these two variables.

3 Data

Seasonally adjusted quarterly data of GDP and unemployment rate for Germany (1991Q1-2017Q4) and Spain (1995Q1-2017Q4) was obtained from Eurostat [10]. GDP is measured in chain linked volumes (reference year 2010, million euro) and the name of the series in Eurostat database is "GDP and main components (output, expenditure and income) [namq_10_gdp]". Unemployment rate includes all ages of unemployed workers (expressed as a number between 0 and 1, not in per cent) and the title of the source data in Eurostat is „Unemployment by sex and age - quarterly average [une_rt_q]". The observable variable y_t corresponds to the natural logarithm of GDP and the variable U_t is simply the unemployment rate.

4 Econometric methodology

Firstly, state space form of the formulated model is described. After that, Kalman filter algorithm is employed in order to construct likelihood function which is then maximized by standard numerical optimization routines.

4.1 State Space Form

Transition and measurement equations of the state space form are given as follows:

$$\mathbf{x_t} = \mathbf{A} \cdot \mathbf{x_{t-1}} + \mathbf{u_t}, \tag{8}$$

$$\mathbf{z_t} = \mathbf{D} \cdot \mathbf{x_t} + \mathbf{v_t}, \tag{9}$$

where $\mathbf{x_t} = \begin{bmatrix} n_t & x_t & x_{t-1} & x_{t-2} & g_t & L_t \end{bmatrix}'$, $\mathbf{z_t} = \begin{bmatrix} y_t & U_t \end{bmatrix}'$,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & \phi_1 & \phi_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & 0 & 1 \end{bmatrix},$$

$\mathbf{u_t} = \begin{bmatrix} v_t & e_t & 0 & 0 & w_t & \varepsilon_t \end{bmatrix}'$, $\mathbf{v_t} = \begin{bmatrix} 0 & \eta_t \end{bmatrix}'$.

Unobserved state vector $\mathbf{x_t}$ can be estimated by Kalman filter once a model is expressed in state space form. However, this would necessitate the knowledge of the matrices $\mathbf{A}$, $\mathbf{D}$, $\mathbf{\Sigma_{uu}} \equiv \mathrm{E}(\mathbf{u_t u_t}')$, $\mathbf{\Sigma_{vv}} \equiv \mathrm{E}(\mathbf{v_t v_t}')$ which are not known in our case because of the unknown parameters $\boldsymbol{\theta} = (\phi_1, \phi_2, \alpha_0, \alpha_1, \alpha_2, \sigma_v, \sigma_w, \sigma_e, \sigma_\varepsilon, \sigma_\eta)$. Therefore, the parameters $\boldsymbol{\theta}$ has to be estimated prior to estimation of the state vector $\mathbf{x_t}$. The vector $\boldsymbol{\theta}$ is econometrically estimated by maximizing likelihood function as Kalman filter can be utilized to calculate likelihood function as well. Kalman filter is summarized in section 4.2. Calculation of the likelihood function is described in chapter 4.3.

4.2 Kalman Filter

Estimation of unobserved state vectors $\mathbf{x_{t+1}}$, $\mathbf{x_t}$ using data up to time t are denoted $\mathbf{x_{t+1|t}}$ and $\mathbf{x_{t|t}}$. Corresponding prediction and estimation error covariance matrices are denoted $\mathbf{P_{t+1|t}} \equiv E\left[\left(\mathbf{x_{t+1}} - \mathbf{x_{t+1|t}}\right) \cdot \left(\mathbf{x_{t+1}} - \mathbf{x_{t+1|t}}\right)' | \mathbf{\Omega_t} \right]$ and $\mathbf{P_{t|t}} \equiv E\left[\left(\mathbf{x_t} - \mathbf{x_{t|t}}\right) \cdot \left(\mathbf{x_t} - \mathbf{x_{t|t}}\right)' | \mathbf{\Omega_t} \right]$, where $\mathbf{\Omega_t} \equiv (\mathbf{z_1}, ..., \mathbf{z_t}, \mathbf{A}, \mathbf{D}, \mathbf{\Sigma_{uu}}, \mathbf{\Sigma_{vv}})$ is information available in time t. Kalman filter calculates these quantities recursively according to

$$\mathbf{x_{t+1|t}} = \mathbf{A} \cdot \mathbf{x_{t|t}}, \tag{10}$$

$$\mathbf{x_{t|t}} = \mathbf{x_{t|t-1}} + \mathbf{K_t} \cdot \left(\mathbf{z_t} - \mathbf{D} \cdot \mathbf{x_{t|t-1}}\right), \tag{11}$$

$$\mathbf{P_{t+1|t}} = \mathbf{A P_{t|t} A}' + \mathbf{\Sigma_{uu}}, \tag{12}$$

$$\mathbf{P_{t|t}} = \mathbf{P_{t|t-1}} - \mathbf{P_{t|t-1}} \mathbf{D}' \left(\mathbf{D P_{t|t-1} D}' + \mathbf{\Sigma_{vv}}\right)^{-1} \mathbf{D P_{t|t-1}}, \tag{13}$$

where $\mathbf{K_t} = \mathbf{P_{t|t-1}} \mathbf{D}' \left(\mathbf{D P_{t|t-1} D}' + \mathbf{\Sigma_{vv}}\right)^{-1}$.

4.3 Likelihood Function

Parameters of the model $\boldsymbol{\theta} = (\phi_1, \phi_2, \alpha_0, \alpha_1, \alpha_2, \sigma_v, \sigma_w, \sigma_e, \sigma_\varepsilon, \sigma_\eta)$ are estimated by maximizing likelihood function $L(\boldsymbol{\theta} | \mathbf{z_1}, ..., \mathbf{z_T})$ which is defined and calculated as follows

$$L(\boldsymbol{\theta} | \mathbf{z_1}, ..., \mathbf{z_T}) \equiv f(\mathbf{z_1}, ..., \mathbf{z_T} | \boldsymbol{\theta}) = \prod_{t=1}^{T} f(\mathbf{z_t} | \mathbf{z_1}, ..., \mathbf{z_{t-1}}, \boldsymbol{\theta}), \tag{14}$$

where $f(\mathbf{z_1},...,\mathbf{z_T}|\boldsymbol{\theta})$ is multivariate density of observed data $\mathbf{z_1},...,\mathbf{z_T}$ given the vector of parameters $\boldsymbol{\theta}$. Conditional densities $f(\mathbf{z_t}|\mathbf{z_1},...,\mathbf{z_{t-1}},\boldsymbol{\theta})$ are multivariate normal densities in our case which is a consequence of normally distributed random errors of the formulated model (1)-(7). The mean and covariance of this conditional distribution are obtained directly from the Kalman filter as follows:

$$\mathbf{z_{t|t-1}} \equiv E\left(\mathbf{z_{t|t-1}}|\boldsymbol{\Omega}_{\mathbf{t-1}}\right) = \mathbf{D}\cdot\mathbf{x_{t|t-1}}\,, \tag{15}$$

$$\mathbf{F_{t|t-1}} \equiv E\left[\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)\cdot\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)'|\boldsymbol{\Omega}_{\mathbf{t-1}}\right] = \mathbf{DP_{t|t-1}D}'+\boldsymbol{\Sigma}_{\mathbf{vv}}\,. \tag{16}$$

The likelihood function (14) can be therefore expressed explicitly in the following manner:

$$L(\boldsymbol{\theta}|\mathbf{z_1},...,\mathbf{z_T}) = \prod_{t=1}^{T}\left[\frac{1}{(2\pi)^{k/2}\left|\mathbf{F_{t|t-1}}\right|}\exp\left(-\frac{1}{2}\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)'\mathbf{F}_{\mathbf{t|t-1}}^{-1}\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)\right)\right], \tag{17}$$

or

$$\ln\left[L(\boldsymbol{\theta}|\mathbf{z_1},...,\mathbf{z_T})\right] = -\frac{T\cdot k}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left|\mathbf{F_{t|t-1}}\right|+\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)'\mathbf{F}_{\mathbf{t|t-1}}^{-1}\left(\mathbf{z_t}-\mathbf{z_{t|t-1}}\right)\right], \tag{18}$$

where k is the number of observed variables.

The function (18) was maximized numerically using standard numerical optimization procedures implemented in Matlab in order to find maximum likelihood estimate $\hat{\boldsymbol{\theta}}$.

5 Estimation results

Estimation results are summarized in the following table 1. Standard errors are indicated in parentheses below estimated coefficients. Standard deviations of random shocks are multiplied by 100 in order to obtain more readable results.[2]

	output decomposition					unemployment decomposition				
	$\hat{\phi}_1$	$\hat{\phi}_2$	$100\hat{\sigma}_v$	$100\hat{\sigma}_w$	$100\hat{\sigma}_e$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$100\hat{\sigma}_\varepsilon$	$100\hat{\sigma}_\eta$
Germany	1.78 (0.05)	-0.79 (0.05)	0.64 (0.05)	0.08 (0.04)	0.35 (0.06)	-0.17 (0.06)	-0.24 (0.10)	-0.03 (0.05)	0.06 (0.05)	0.01 (0.08)
Spain	1.62 (0.09)	-0.64 (0.09)	0.11 (0.04)	0.23 (0.03)	0.09 (0.02)	-3.43 (1.39)	-2.18 (0.98)	-0.00 (0.60)	0.09 (0.60)	0.08 (0.20)

Table 1 Estimation results of the bivariate unobserved components model (1)–(7) for Germany (1991Q1-2017Q4) and Spain (1995Q1-2017Q4)

Parameters ϕ_1 and ϕ_2 are statistically significant (at standard 5 % level) and the sum $\hat{\phi}_1+\hat{\phi}_2$ is close to one for both countries which means the output gap x_t is highly persistent. Clark [9] reported estimates $\hat{\phi}_1 = 1.47$ and $\hat{\phi}_2 = -0.59$ for the US economy (1947Q1-1986Q2). The sum of these parameters for US economy $\hat{\phi}_1+\hat{\phi}_2 = 0.88$ is not as close to 1 as in Germany ($\hat{\phi}_1+\hat{\phi}_2 = 0.99$) and Spain ($\hat{\phi}_1+\hat{\phi}_2 = 0.98$). Output gap in Germany and Spain is thus remarkably more persistent than the corresponding gap in US economy.

As far as the relation between output gap and cyclical unemployment rate is concerned, parameters $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ have expected negative sign. Parameters α_0 and α_1 turned out to be statistically significant (at standard 5 % level) which is not the case for α_2. Thus, only current and one quarter lagged output gap has significant influence on unemployment rate gap. These results are in line with Clark's [9] formulation of the Okun's law as he assumed that only current output gap x_t and output gap lagged one period x_{t-1} have influence on unemployment gap C_t. His estimates (in our notation) for US economy are $\hat{\alpha}_0 = -0.33$ and $\hat{\alpha}_1 = -0.18$. The overall effect

[2] Alternatively, the data for log of GDP and unemployment rate could be multiplied by 100 in order to increase estimated standard errors of random shocks 100 times.

of output gap on unemployment gap according to the Clark's estimate is $\hat{\alpha}_0 + \hat{\alpha}_1 = -0.51$. His result is quite close to the estimated overall effect for Germany ($\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 = -0.44$). These findings are in line with Mankiw [16] who posits that (for US economy) a one percent deviation of output from potential causes an opposite change in unemployment rate of half a percentage point. This assertion tells practically the same as the Clark's [9] estimate $\hat{\alpha}_0 + \hat{\alpha}_1 = -0.51$ and also corresponds closely to our estimate for Germany. Nonetheless, estimation results for Spain ($\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 = -5.61$) indicate much stronger influence of the output gap on cyclical properties of unemployment rate which can be explained by the prevalence of temporary employment contracts in Spain.

Boďa et al. [6] estimates Okun's law for Visegrad group countries by applying nonlinear autoregressive distributed lag (NARDL) methodology using quarterly data from 1998 Q1 to 2014 Q2. Their long-run multiplier is comparable to our estimate $\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2$ and they report the following values – Czech Republic (-0.12), Slovakia (-0.20), Hungary (-0.06) and Poland (-0.08). Their estimates are thus even smaller than our estimate for Germany or Clark's estimate for US economy. This suggests that there are considerable differences regarding the strength of the Okun's law among individual European countries. This is an interesting finding as lots of empirical research estimate Okun's law for the whole Europe using aggregated data (Orlandi, Pichelman [19], Berger [3], Berger, Everaert [4]). There are also papers estimating Okun's law by applying disaggregated approach using data for individual economies, but assuming that Okun's law coefficients are the same for all countries (Novák, Darmo [17]).

Novák, Darmo [17] estimated Okun's law in its difference form and found structural breaks. Nonetheless, Brůha, Polanský [7] argue that instability is often found when using a difference form of Okun's law and found evidence that the relationship is stable when cyclical components of output and unemployment are used instead of differences. Similarly, Ball et al. [2] estimated Okun's law in its gap form and state that Okun's law is a strong and stable relationship in most countries, one that did not change substantially during the Great Recession. As was already discussed, Ball et al. [2] also found that Okun's law coefficient varies substantially across countries, which is confirmed by the results presented in this paper as well. Therefore, the assumption of constancy of this parameter across individual economies introduced by Novák, Darmo [17] is not satisfied, which makes their results less viable.

6 Conclusion

Parameters of the bivariate unobserved components Clark's [9] model were estimated for Germany and Spain with the main goal of discussing the relationship between output gap and cyclical unemployment rate known as Okun's law. Econometric estimation revealed that output has a highly persistent stationary cyclical component for both Germany and Spain. Comparison with Clark's [9] estimates for US economy confirmed that persistence of the gap in both these countries is much higher than in US.

Parameter estimation of the Okun's law describing relation of the output gap to cyclical unemployment rate showed that all these coefficients had expected negative sign. This negative relation confirms an empirical validity of the Okun's law in its gap form which implies that cyclical properties of unemployment rate are driven mainly by cyclical component of GDP. Estimated Okun's law coefficients for Germany turned out to be in line with Clark's [9] estimated overall effect (–0.51) of the output gap on cyclical unemployment rate for US economy positing that a one percent deviation of output from potential causes an opposite change in unemployment rate of half a percentage point. Nonetheless, this is not the case for Spain where the estimated overall effect turned out to be approximately 10-times higher.

Findings regarding substantial heterogeneity in the strength of the Okun's law across individual European countries emphasize importance of analyzing disaggregated data as many studies use only aggregated data for the whole Europe (Azevedo et al. [1], Berger [3], Berger, Everaert [4], Bernhofer et al. [5], Chen, Mills [8], Galati et al. [11], Orlandi, Pichelmann [19] or Proietti [20]). Possible structural instability of the Okun's law advocated by Novák, Darmo [17] has not been detected and previous results regarding structural stability of the Okun's law published by Brůha, Polanský [7] or Ball et al. [2] have been confirmed.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/21/2018.

References

[1] Azevedo, J., Koopman, S., Rua, A. Tracking Growth and the Business Cycle: A Stochastic Common Cycle Model for the Euro Area. *Tinbergen Institute Discussion Paper No. TI 2003-069/4*, 2003.

[2] Ball, L., Leigh, D., Loungani, P. Okun's Law: Fit at 50? *Journal of Money, Credit and Banking*, **49**(7), 2017, 1413–1441.

[3] Berger, T. Estimating Europe's Natural Rates. *Empirical Economics*, **40**(2), 2011, 521-536.

[4] Berger, T., Everaert, G. Unemployment Persistence and the NAIRU: A Bayesian Approach. *Scottish Journal of Political Economy*, **55**(3), 2008, 281-299.

[5] Bernhofer, D., Amador, O., Gächter, M., Sindermann, F. Finance, Potential Output and the Business Cycle: Empirical Evidence from Selected Advanced and CESEE Economies. *Focus on European Economic Integration*, **2**, 2014, 52-75.

[6] Boďa, M., Medveďová, P., Považanová, M. (A)symmetry in Okun's Law in the Visegrad Group Countries (in Slovak). *Politická ekonomie*, **63**(6), 2015, 741-758.

[7] Brůha, J., Polanský, J. Empirical analysis of labor markets over business cycles: an international comparison. Working Papers 2015/15, Czech National Bank, Research Department, 2015

[8] Chen, X., Mills, T. C. Measuring the Euro Area Output Gap Using a Multivariate Unobserved Components Model Containing Phase Shifts. *Empirical Economics*, **43**(2), 2012, 671-692.

[9] Clark, P. K. Trend Reversion in Real Output and Unemployment. *Journal of Econometrics*, **40**, 1989, 15-32.

[10] Eurostat. Database–Eurostat, [online.] [Accessed 1 May 2018.] <http://ec.europa.eu/eurostat>

[11] Galati, G., Hindrayanto, I., Koopman, S., Vlekke, M. Measuring Financial Cycles with a Model-Based Filter: Empirical Evidence for the United States and the Euro Area. *DNB Working Paper No. 495*, 2016.

[12] Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. 1st Ed. Cambridge: Cambridge University Press, 1989.

[13] Harvey, A. C., Jaeger, A. Detrending, Stylized Facts and the Business Cycle. *Journal of Applied Econometrics*, **8**(3), 1993, 231-247.

[14] Hodrick, R. J., Prescott, E. C. Postwar U.S. business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, **29**(1), 1997, 1-16.

[15] Kim, C. J., Nelson, C. R.: *State Space Models with Regime Switching*. MIT Press, Cambridge, 1999.

[16] Mankiw, G. *Principles of Macroeconomics*, 6th edition, 2012.

[17] Novák, M., Darmo, L. Okun's Law over the Business Cycle: Does It Change in the EU Countries after the Financial Crisis? *Prague Economic Papers*, **28**(2), 2019, 235-254.

[18] Okun, A. Potential GNP: Its Measurement and Significance. *American Statistical Association: Proceeding of the Business and Economics Statistics Section*, 1962, 98-104.

[19] Orlandi, F., Pichelmann, K. Disentangling Trend and Cycle in the EUR-11 Unemployment Series: An Unobserved Component Modelling Approach. *Brussels: European commision*, 2000.

[20] Proietti, T. State Space Decomposition under the Hypothesis of Non-Zero Correlation between Trend and Cycle with an Application to the Euro-zone. In Monographs of Official Statistics, Papers and Proceedings of the Third Colloquium on Modern Tools for Business Cycle Analysis, L. L. Mazzi et H. Savio (Eds.), *Research in Official Statistics*, 292-325, Eurostat, Luxembourg, 2004.

Methodological framework for 'passive' and 'active' approach to the investment portfolio making

Adam Borovička[1]

Abstract. The (optimization) approach to an investment portfolio making can be classified as 'passive' and 'active'. The 'passive' form is methodologically represented by using a well-known mean-variance model including standardly two criteria – return and risk. The 'active' approach provides a more complex portfolio selection process. Then other evaluating investment criteria (cost, market mood, etc.) are considered. Additional investor's preferences can be also formulated, i.e. an explicit specification of the criteria importance or a determination of the goal (required) values of the portfolio characteristics. All these aspects can be satisfactorily taken into account through an interactive multiple objective programming method. This method is developed on the basis of a well-known STEP method. Advantages and pitfalls of the proposed methodological concept are empirically demonstrated on real-life making a portfolio of open unit trusts offered by Česká spořitelna. Both portfolio making processes are empirically analyzed and compared.

Keywords: 'Active' approach, mean-variance model, multi-objective, open unit trust, 'passive' approach, portfolio making.

JEL Classification: C61, G11
AMS Classification: 90C90

1 Introduction

Most people are not very familiar with the world of finance although they are investing on the capital market. Their main evaluating criterion is return that is occasionally linked to the investment risk (related to the possible loss). Other characteristics, as well as additional investor's preferences are not usually specified. The investment decision is usually made in close cooperation with the investment counsel. This way can be called '*passive'* approach. Typical opposite approach may be marked as '*active'*. Such an investor takes into account other decision criteria (fees as the cost connected to the investment, investment locality, or mood on the capital market). Further, s/he usually explicitly expresses a potentially diverse importance of all criteria. S/he can specifies some goal (required) values of the selected portfolio characteristics. This investor usually enables to make the investment analysis without external advisory person. The 'active' strategy can be also applied with an investment council when not too many self-collected input information (data) from the capital market are available.

The study of the last few decades confirms that the optimization principle is a effective means for an investment portfolio making. The foundation was laid by H. Markowitz with his mean-variance model [7, 8]. This concept is superbly applicable for the 'passive' portfolio making. It provides the 'effective' portfolios from the perspective of two most important characteristics – return and risk. Moreover, this model is well understandable. Its application is user-friendly. However, this model is limited for the 'active' investment approach because other criteria or additional investor's preferences cannot be considered. The 'active' approach requires a portfolio making procedure that allows to find 'tailored-made' solution. Therefore, an interactive multiple objective programming method is designed. Although this is a non-traditional tool for using on capital market, it proves to be very supporting for the investment portfolio making, especially for its 'active' form. Such a method must be able to take into account all the necessary criteria and investor's preference about their importance or required levels. The integrated interactive procedure helps to finding a tailored-made portfolio through its gradual changes based on continually expressed preferences (requirements) by the investor. A user-friendly method meeting these assumptions is STEP method [3]. However, some partial principle in this algorithm is not completely perfect. To eliminate these smaller drawbacks, proper modifications are proposed.

The practical part of this article provides an empirical study of both investment approaches using the proposed methodological concepts. The portfolios are made from the open unit trusts offered by Česká spořitelna that becomes still more popular investment instruments in recent time. The real-life portfolio making cases demonstrate the abilities and skills of both designed concepts for 'passive' and 'active' approach to the investing.

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

In summary, the main aim, or mission of this article is to propose a methodological framework for a portfolio making based on the effective optimization principle. To cover the most investment cases, the optimization method is designed for 'passive' and 'active' approach to the investing. Subsequent aim is to demonstrate the applicability of the proposed methodological concept on a real-life portfolio making process.

The structure of the paper is as follows. The introduction is followed by Section 2 proposing a methodological framework for 'passive' and 'active' investment portfolio making. Section 3 offers an empirical analysis of both approaches on the Czech capital market with open unit trusts. In the end (in Section 4), the main contributions of the article are summarized and some incentive ideas for future research are outlined.

2 Methodological framework for the investment portfolio making

The methodology for a portfolio making includes the separate optimization concepts for both investment approaches. The portfolio for a 'passive' investment strategy is made by a 'classic' mean-variance model from H. Markowitz [7, 8]. The 'active' approach emphasizes a more complex multi-criteria perspective to the investing with a gradual interactive search of the most suitable portfolio composition. Therefore, the interactive multiple objective programming method (revised STEP Method) is designed.

2.1 'Passive' approach

The passive approach can be classified as the most widespread. Such a (potential) investor usually doesn't understand investing too much. All steps are taken by this investor through a bank employee, or investment counsel. His/her investment strategy is rather long-term, without any dramatic changes in the investment portfolio in time. Virtually the only evaluating criteria for investment are its (expected) return and risk connected with a potential loss. Other characteristics are negligible.

Then the most suitable optimization tool for a portfolio selection under these conditions is Markowitz mean-variance model [7, 8]. The investor obtains a portfolio composition with an 'effective' combination of its (expected) return and risk. This approach is user-friendly. No additional preferences (about criteria importance, goal values, etc.) are required. It could be comprehensible for a 'passive'investor or his/her investment counsel. As this investor tends to be more conservative, the mean-risk model can be formulated in the minimizing form. This model diversifies the risk under some lower level of the investment return. Then the model is formulated as follows

$$\begin{aligned} \min \quad & r_p(\mathbf{x}) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij}} \\ \text{s.t.} \quad & \sum_{i=1}^{n} er_i x_i \geq er^* \\ & \mathbf{x} \in X \end{aligned} \quad , \tag{1}$$

where $X = \left\{ \mathbf{x} \in R^n ; \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0,\ i = 1,2,...,n \right\}$. The vector $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ includes the variables x_i, or x_j $(i, j = 1,2,...,n)$ representing a share of the *i*-th, or *j*-th asset (open unit trust in this research) in the portfolio, er_i $(i = 1,2,...,n)$ denotes an (expected) return of the *i*-th asset, σ_{ij} $(i, j = 1,2,...,n)$ marks a covariance (variance) of return of the *i*-th and *j*-th asset. Lower limit of the portfolio return er^* can be derived from the best possible return under the conditions created set *X* through the one-objective optimization model (see (3) below).

2.2 'Active' approach

The second most common approach to investing can be called 'active'. Such an investor has a knowledge about the investing. S/he orients in the world of finance and investment. S/he is open to new things in this field. S/he has a clear idea of his/her investment. Besides return and risk, s/he takes into account other criteria (cost, investment locality, etc.). Moreover, the investor has clarified the preferences regarding the importance of these criteria. S/he has a fairly good idea of the value of investment portfolio characteristics. This investor is usually bolder. S/he is able to take a greater risk to potentially gain a higher return. However, s/he may also be risk-averse.

To consider all aspects of this approach, some interactive multiple objective methods can be a very good (technical) supporting tool for an investment portfolio selection. These methods allow a sequential creation of the portfolio according to the investor's preferences which may be possibly changed during the process. These methods

can be classified through the form of an expression of substitution. This substitution can be explicit, i.e. a quantitative substitution rate between the values of two objective functions [4]. The father of these approaches is GDF method [5]. The most important problem may be a determination of the substitution rates. This could be difficult even for the experienced investor. Further, many of these methods work with a utility function, e.g. [9]. This function may not be explicitly known which often makes finding a solution more difficult. A more suitable approach seems to be the implicit expression of the substitution. Then two main groups of the objective functions are determined – objective functions with unacceptable value and objective functions with an acceptable deterioration in their values. The part of these approaches is also based on the utility function principle in which I see the weaknesses. A construction of the utility function can be complicated that potentially can make the process unnecessarily difficult. The second evaluating optimization principle can be based on the minimization of distance from the ideal solution. In this group, I am searching the method with implicit substitution rates and understandable (user-friendly) algorithm for a potentially wider range of users (investors). After a shorter research, one of the oldest interactive methods, STEP Method, seemed to be a suitable candidate. However, this method also shows some drawback. The weights of criteria are determined by the algorithm. This fact is unfavorable because the investor wants to set the criteria weights on his own preferences. So, the algorithm of the original method must be slightly modified (generalized) as follows:

Step 1: Indicate the set of minimizing, or maximizing objective functions $\left\{ f_j(\mathbf{x}), j = 1,2,...,k_1 \right\}$, or $\left\{ f_j(\mathbf{x}), j = k_1+1, k_1+2,...,k_2 \right\}$, the set X^1 containing all conditions (the same as X in (1), or extended by possibly others) and the vector of criteria weights $\mathbf{w} = (w_1, w_2,..., w_{k_1}, w_{k_1+1}, w_{k_1+2},..., w_{k_2})^T$. Compared to the original algorithm, the weights of criteria are determined according to the decision maker (investor) preferences (e.g.) via a well-known scoring method. Then the ideal value of each objective function is found on the set X^1. Let denote the optimal value of j-th minimizing, or maximizing objective function as $f_j^* = f_j(\mathbf{x}_\mathbf{j}^*)$, when it holds for $j = 1,2,...,k_1$, or $j = k_1+1, k_1+2,...,k_2$ the following

$$\mathbf{x}_\mathbf{j}^* = \arg\min_{\mathbf{x} \in X^1} f_j(\mathbf{x}) \text{, or } \quad \mathbf{x}_\mathbf{j}^* = \arg\max_{\mathbf{x} \in X^1} f_j(\mathbf{x}) . \tag{2}$$

Step 2: Now, the following mathematical model is solved

$$\begin{aligned} \min \quad & D \\ \text{s.t.} \quad & w_j \frac{\left| f_j(\mathbf{x}) - f_j^* \right|}{f_j^*} \le D \quad j = 1,2,...,k_1, k_1+1, k_1+2,...,k_2 . \\ & \mathbf{x} \in X^1 \end{aligned} \tag{3}$$

This model minimizes the maximal weighted standardized (unlike the original concept) deviation (over all objective functions) from the ideal alternative (solution). The optimal solution of (3) is the vector $\mathbf{x^1}$ with the values of objective functions $f_j^1 = f_j(\mathbf{x^1}), j = 1,2,...,k_1, k_1+1,...,k_2$. If the values of all objective functions are acceptable by a decision maker (investor), the compromise solution $\mathbf{x^1}$ (portfolio composition) is found. If the values of all objective functions are not acceptable, the algorithm also ends because the values of all objective functions cannot be simultaneously improved. Solution $\mathbf{x^1}$ is non-dominated. If the values of some objective functions are acceptable and some unacceptable, the interactive procedure is started in terms of the following step.

Step 3: Let denote the sets containing the indices of minimizing, or maximizing objective function with a non-acceptable values as I_{min}^{na}, or $I_{\max}^{na}$. Further, the sets of the indices of minimizing, or maximizing objective functions with acceptable values are denoted as I_{min}^{a}, or I_{max}^{a}. Finally, the sets of the indices of minimizing, or maximizing objective functions with a possible deterioration of their values are marked as I_{min}^{ad}, or I_{max}^{ad}. Then the set of feasible solutions is extended by the following set X^2

$$X^2 = \left\{ \begin{array}{llll} f_j(\mathbf{x}) \le f_j^1 - \varepsilon & j \in I_{min}^{na}, & f_j(\mathbf{x}) \ge f_j^1 + \varepsilon & j \in I_{max}^{na} \\ f_j(\mathbf{x}) \le f_j^1 & j \in I_{min}^{a}, & f_j(\mathbf{x}) \ge f_j^1 & j \in I_{max}^{a} \\ f_j(\mathbf{x}) \le f_j^1 + \Delta_j & j \in I_{min}^{ad}, & f_j(\mathbf{x}) \ge f_j^1 - \Delta_j & j \in I_{max}^{ad} \end{array} \right\}, \tag{4}$$

where ε is a small positive number close to zero (e.g. 10^{-4}) to ensure not sharp inequalities (compared to the original formulation). Δ_j expresses a possible deterioration of the value of the *j*-th objective function. Finally, model (3) is adequately modified as follows

$$\begin{aligned} \min \quad & D \\ \text{s.t.} \quad & w_j \frac{\left| f_j(\mathbf{x}) - f_j^* \right|}{f_j^*} \le D \quad j = 1, 2, ..., k_1, k_1 + 1, k_1 + 2, ..., k_2 \\ & w_j = 0 \quad j \in I_{min}^{ad} \cup I_{max}^{ad} \\ & \mathbf{x} \in X^1 \cup X^2 \end{aligned} . \tag{5}$$

The optimal solution of model (5) $\mathbf{x}^2$ is found. If this solution is still nonacceptable, Step 3 is repeated. If the solution is acceptable, the compromise solution is found. The algorithm is terminated.

Models (2) and (3) have a solution because of a limited (non-empty) set of feasible solutions. Model (5) is also solvable if the proposed changes of the objective functions' values are attainable. Due to the risk (diversification) function, the models are nonlinear. Then only local optimum can be mostly found.

3 Empirical study of both investment approaches

The empirical study of both investment approaches is performed on making a portfolio from open unit trusts. These investment instruments are increasingly popular in the Czech Republic [1]. It is also suitable for 'smaller' investors within a middle or long-time investment horizon. The open unit trusts offered by Česká spořitelna are selected. It is one of the biggest providers of this product in the Czech Republic [1]. Moreover, the author has gained some experiences with the investing in these funds. From the perspectives of both approaches, three main characteristics are explicitly observed. The (expected) return is calculated as monthly average. The risk is measure by a variance of returns. Both characteristics are calculated form the historical prices from the period 2006-2019. The investment in open unit trusts is rather longer-time investment, therefore the selected period should reflect a longer-time development. A sufficient number of observations (for risk calculation) is just ensured by this time period and also monthly frequency of returns. The cost is represented by the input fee (a percentage from the invested amount). Other fees (management, license, etc.) are not explicitly considered, because they are included in the price, or return of the funds.

The data (in percentage) about 22 open unit trusts, offered by Česká spořitelna in the period 2006-2019, are shown in the following table (Table 1). The prices and fees are collected from the web side of Investment Centre [6]. The returns and their variances are calculated from the funds' prices.

Fund	Return	Risk	Cost	Fund	Return	Risk	Cost
ConRF	0.0136	0.2489	1	CzoSB	0.1727	1.0607	2.5
ConM	0.0995	1.4204	1	CFoCB	0.0583	0.2341	0.5
BalM	0.1518	4.1575	1.5	BonEHY	0.3817	11.1770	3.5
DynM	0.1537	8.7946	1.5	SporT	0.0095	52.8181	3
StoM	0.2134	15.1478	3	GlobalS	0.3766	16.2185	3
SporIn	0.0292	0.0892	0.3	TopS	0.8961	50.0959	3
SporB	0.1959	1.0238	1	StockJ	0.0970	22.3363	4
TreB	0.0560	3.6953	1	StockI	0.2014	94.1828	4
CorBF	0.2105	6.5270	1	StockG	0.2880	21.2174	4
HighY	0.2758	9.7080	1	StockEP	0.2668	29.2766	4
PorBE	0.2453	4.0442	3.5	StockEE	-0.2049	44.9877	4

Table 1 Open unit trusts data

In Table 1, we can see five balanced funds (*Controlled Returns Fund* – abbreviated as ConRF, *Conservative Mix* – ConM, *Balanced Mix* – BalM, *Dynamic Mix* – DynM, *Stock Mix* - StoM), nine bond funds (*Sporoinvest* – SporIn, Sporobond – SporB, Trendbond – TreB, *Corporate Bonds Fund* – CorBF, *High Yield* – HighY, *Portfolio Bond Europe* – PorBE, *Czech Fund of State Bonds* – CzoSB, *Czech Fund of Corporate Bonds*- CFoCB, *Bond Europe High Yield* – BonEHY) and eight equity funds (Sporotrend – SporT, *Global Stock* – GlobalS, *Tops Stocks* – TopS,

Stock Japan – StockJ, *Stock Istanbul* – StockI, *Stock Global* – StockG, *Stock Europe Property* – StockEP, *Stock Europe Emerging* – StockEE).

For a 'passive' approach, a mean-variance model (1) is formulated. It is obvious from Table 1, that the biggest possible return of the portfolio is 0,8961%, the lowest is -0.205%. A mean-variance model for a various minimum required expected return is formulated and solved in Lingo optimization software. The efficient combinations of risk and return (as efficient frontier) are shown in the following figure (Figure 1).

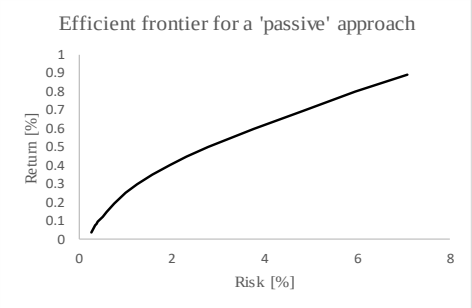


Figure 1 Efficient frontier for a 'passive' approach

Without additional investor's preferences, the portfolio for a required minimum return with the smallest possible risk can be revealed. For instance, the required return equal to half a maximum value of possible return generates the lowest possible risk at the level of 2.33%. The value of an unmonitored characteristic cost is 3.31%. Portfolio has the following form: 65.23% *ESPA Portfolio Bond Europe*, 31.52% *Top Stocks* and 3.25% *ESPA Czech Fund of State Bonds*. The combination of ESPA Portfolio Bond Europe and Top Stocks funds is not surprise, because the covariances of their returns is negative. Moreover, they are able to ensure the required return. None of balanced funds is in the portfolio because of their too low returns. By the way, these funds have rather higher cost. Then it is expectable that a consideration of this characteristic would cause a different portfolio composition.

As mentioned above the 'active' investor is consider as rather "return-seeking". S/he is able to undergo a higher level of risk to potentially obtain some additional return. This makes it more appealing to the cost of investment. Further, s/he can provide other initiative requirements for a portfolio composition. However, these additional conditions are not included in this analysis to make the results comparable to the situation of a 'passive' approach. Based on the investor's preferences, the weights are determined (via a scoring method) as follows: return – 0.6, risk – 0.25, cost – 0.15. The initial portfolio for an 'active' investor is found through the following model (as (5))

$$\begin{aligned} &\min \quad D \\ &\text{s.t.} \quad 0.6\frac{\left|\sum_{i=1}^{22} er_i x_i - 0.8961\right|}{0.8961} \le D \quad 0.25\frac{\left|\sqrt{\sum_{i=1}^{22}\sum_{j=1}^{22} x_i x_j \sigma_{ij}} - 0.2768\right|}{0.2768} \le D \quad 0.15\frac{\left|\sum_{i=1}^{22} cost_i x_i - 0.3\right|}{0.3} \le D, \\ &\quad \mathbf{x} \in X^1 \end{aligned} \tag{6}$$

where $X^1 = \left\{ \sum_{i=1}^{22} x_i = 1;\ x_i \ge 0,\ i = 1,2,...,22 \right\}$, x_i is a share of the *i*-th open unit trust (in the order from the left column in the Table 1) and $cost_i$ marks the cost of the *i*-th open unit trust. The best (optimal) values of return, risk and cost is found on the set X^1 (see model (2)). The result of (6), received by Lingo software, represents the initial portfolio in the following form: 51.31% *Sporobond*, 12.19% *ESPA Portfolio Bond Europe*, 31.68% *ESPA Czech Fund of Corporate Bonds* and 4.82% *Top Stocks*. Its expected monthly return is 0.1921%, risk 0.7987% and cost 1.2428%. The covariances of returns of ESPA Portfolio Bond Europe fund are mostly negative over all open unit trusts which is unique compared to the others. Its participation is confirmed by a solid return. The fund with the highest return Top Stocks cannot be missed in the portfolio. Its bigger share is limited by its strong positive correlation of its return with other funds's return, which weakens its 'diversification skills'. For the return mentioned above, the mean-variance model identifies the portfolio with only slightly less risk (approximately 0.03% difference). However, the cost is significantly lower (approximately 0.5% difference). This fact can be very important for the investor strictly focused on the return. It is obvious, the interactive multiple objective programming method is very beneficial for the 'active' approach. Of course, the investor can require some improvement of values of the selected objectives. Since s/he mainly focuses on return, s/he requires a return increase at the same level of risk. Less significant cost may deteriorate by 0.2%. Then the model (5) in the following form is solved

$$\begin{aligned} &\min \quad D \\ &\text{s.t.} \quad 0.6\frac{\left|\sum_{i=1}^{22} er_i x_i - 0.8961\right|}{0.8961} \le D \quad 0.25\frac{\left|\sqrt{\sum_{i=1}^{22}\sum_{j=1}^{22} x_i x_j \sigma_{ij}} - 0.2768\right|}{0.2768} \le D \quad 0\frac{\left|\sum_{i=1}^{22} cost_i x_i - 0.3\right|}{0.3} \le D, \\ &\quad \mathbf{x} \in X^1 \cup X^2 \end{aligned} \tag{7}$$

where $X^2 = \{ f_1(\mathbf{x}) \ge 0.1921 + 0.0001,\ f_2(\mathbf{x}) \le 0.7987,\ f_3(\mathbf{x}) \le 1.2428 + 0.2 \}$. The portfolio composition remains the same. The significant change is an increase of the share of ESPA Portfolio Bond Europe and reduction of the Sporobond share. If a higher deterioration of cost is accepted, a 'mean-risk' effective portfolio will be revealed. Its characteristics are: return – 0.1999%, risk – 0.793%, cost – 1.7977%. However, a significant increase of cost doesn't produce an adequate rise of the return. The investor would accept the portfolio found by model (7), or the initial version. Finally, let analyze the case with a higher importance of the cost in order to more accentuate a net profit of the investment. For instance, the weight of return is the same, but the weight of risk is only 0.15 and cost weight is 0.25. The portfolio has the following structure: 54.79% *Sporoinvest*, 32.74% *Sporobond* and 12.47% *Top Stocks*. The portfolio characteristics are as follows: return – 0.1918%, risk – 1.147%, cost – 0.8659%. For this level of return, the 'mean-variance' model provides a lower level of risk. On the other hand, it is much more costly. It confirms that the interactive procedure (unlike a mean-variance model) enables to satisfactorily take into account all investor's preferences and requirements for the portfolio.

4 Conclusion

The article provides an optimization methodological framework for 'passive' and 'active' approach to the investing. For the 'passive' strategy, a well-known mean-variance concept is applied. Determination of the effective portfolios from the perspective of two closely monitored characteristics (return and risk) is very beneficial. To include other criteria or requirements for the portfolio, a more complex multi-objective model is developed. Then for such an ‚active' approach, the interactive multiple objective programming method is designed. Both methods and their benefits, or limits (related to the investment approach) are presented on making a portfolio of open unit trusts traded on the Czech market. The empirical analysis proves a strong application ability of the introduced methodological concepts for the specified 'passive' and 'active' approaches to the investment portfolio making. In the future research, to eliminate a (strict) assumption of returns normality, a mean-variance concept for 'passive' approach should be replaced by a mean-semivariance model [2, 8]. Finally, the theoretical and empirical study of various risk measures can be very interesting and beneficial for a decision making on the capital market.

Acknowledgements

The research project was supported by Grant No. F4/66/2019 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] AKAT ČR. (2019). *Annual press conference - presentation*. [online] Available at: file:///C:/Users/adamb/Downloads/tk20190305_prezentace_roku_2018.pdf, [Accessed 30 April 2019].
[2] Ballestero, E. (2007). Mean-semivariance efficient frontier: A downside risk model for portfolio selection. *Applied Mathematical Finance*, *12*, 1–15.
[3] Benayoun, R., de Montgolfier, J., Tergny, J. & Laritchev, O. I. (1973). Linear programming with multiple objective functions: STEP Method (STEM). *Mathematical Programming*, *1*, 366–375.
[4] Fiala, P. (2013). *Models and methods of a decision making* (in Czech). Prague: Oeconomica.
[5] Geoffrion, A. M., Dayer, J. S. & Feinberg, A. (1972). An interactive approach for multi-criterion optimization, with an application to the operation an academic department. *Management Science*, *4*, 357–367.
[6] Investment center. (2018). *Funds*. [online] Available at: https://cz.products.erstegroup.com/Retail/en/index.phtml, [Accessed 25 April 2019].
[7] Markowitz, H. M. (1952). Portfolio Selection. *Journal of Science*, *7*, 77–91.
[8] Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. New York: John Wiley & Sons, Inc.
[9] Zionts, S. & Wallenius, J. (1976). An interactive programming method for solving the multiple criteria problem. *Management Science*, *22*, 652–663.

Detecting hidden patterns of exchange rate returns with recurrence analysis

Jan Kodera[1], Quang Van Tran[1]

Abstract. Theorists as well as practitioners have been trying hard to identify regular patterns in fluctuations of exchange rate returns for a long time. However, their investigation has often been conducted with traditional econometric technique. In the last few decades, non-linear methods of time series analysis have started being employed to examine behavior of economic and financial systems. These techniques have often been so far applied on processes which are considered to be stochastic. In this contribution, first, we use visual recurrence analysis (VRA) to graphically detect patterns and structural changes hidden in financial data. Then these hidden patterns and structures will be quantified by using recurrence quantification analysis (RQA). These two nonlinear time series methods will be applied to determine hidden patterns and structures of fluctuations of exchange rate returns present in series EURCZK and USDCZK in the last 9 years.

Keywords: Nonlinear time series analysis, visual recurrence analysis, recurrence quantification analysis, Czech crown exchange rates.

JEL Classification: C58, F31
AMS Classification: 37M10

1 Introduction

Financial time series have been mostly investigated from the perspective of stochastic processes in order to find some regular patterns in their fluctuations. However, it is well documented that nonlinear deterministic systems can produce dynamics whose projections onto time dimension resemble the trajectories of stochastic development. In the recent past, many non-linear methods of time series analysis have been developed and they have been widely used in many technical disciplines. Later, they have been used for investigating behavior of economic and financial time series. One of those methods is the visual recurrence analysis (VRA) which can be applied to quickly detect patterns and structural changes hidden in data. In this research, the VRA will be employed to examine the hidden patterns and structures of fluctuations of exchange rate returns in series EURCZK and USDCZK exchange rate returns from 2010 – 2019. Besides the qualitative analysis using VRA, the hidden patterns and structures will be quantified by using recurrence quantification analysis (RQA). As the recurrence analysis is performed on the states of a dynamic system in a phase space, the reconstruction of the state space from time series needs to be conducted. In the paper, we briefly describe this reconstruction technique and carry it out from those examined series before VRA and RQA are performed.

2 Visual Recurrence Analysis

The Visual Recurrence Analysis is a new technique proposed by Ekmann et al. [2] which uses recurrence graphics to detect patterns and structural changes hidden in the data. By this way, the dynamics present in time series can be qualitatively studied as the time series can be seen as a realization of interaction among many factors of a highly complex dynamic system. From the mathematical point of view, the recurrence map is the results of the projection of in general an m-dimensional faze-space dynamic system on $\mathbb{R}^2$. A recurrence map is a graphical representation of a square matrix of size $N \times N$, whose elements are either 0 or 1 indicating the recurrence of a state over time. If a state at time i y_i reoccurs at time j, it has value 1 in this matrix, otherwise, it gets 0 and in the recurrence graph each value gets marked by a different color. Formally it can be expressed as follows

$$R_{i,j}(\epsilon) = \Theta\left(\epsilon - \| y_i - y_j \|\right), \text{for } i, j = 1, 2, \ldots, N, \tag{1}$$

where Θ is the Heaviside function, ϵ is a threshold distance, y_i, y_j are states at time i, j respectively, N is the total number points of the time series.
Recurrence graph by definition is symmetric around line 45^o and can have any combination of these structures:

[1] University of Economics in Prague, Faculty of Finances and Accounting, náměstí W. Churchilla 4, Praha 3, 130 00 - Czech Republic, kodera@vse.cz, tran@vse.cz

– individual points which mean that the states in the phase space are unique, the system does not stay on them for a long time or it strongly fluctuates

– diagonal lines, where the length of a diagonal line occur when a part of the trajectory runs parallel to another part of the trajectory, or trajectory returns to the same area in different times. These lines indicate the existence of unstable periodic orbits and they can be a hallmark of determinism. If there are only these diagonal lines on the recursive graph, it is a periodic signal (the distances from one state are repeated periodically). The length of this section is determined by how long the individual parts of the trajectory remain in these areas,

– vertical or horizontal lines, which means that states remain at one point or little change from this point, so the system is trapped at these points.

3 Recurrence Quantification Analysis

Though recurrence plot may be a useful tool for the fast detection of presence of deterministic nonlinearity in time series, it can only provide a qualitative answer. Further, there may be some patterns that are difficult to recognize by visual checking. In order to overcome this problem, Zbilut and Webber [8] and later Marvan et al. [6] propose the so called Recurrence Quantification Analysis (RQA), which quantifies properties of recurrence plot by several statistics. Before we describe these statistics, besides $R_{i,j}$ in (1) let's define some auxiliary quantities. The diagonal line $D_{i,j}$ is defined as

$$D_{i,j} = \begin{cases} 1 \text{ if states } (i-1,j-1),(i,j),(i+1,j+1) \text{ are recurrent} \\ 0 \text{ otherwise} \end{cases}, \tag{2}$$

the vertical line is

$$V_{i,j} = \begin{cases} 1 \text{ if states } (i,j-1),(i,j),(i,j+1) \text{ are recurrent} \\ 0 \text{ otherwise,} \end{cases} \tag{3}$$

the probability of recurrence of a diagonal line of length d is

$$P(d) = \sum_{i,j}^{N} (1 - R_{i-1,j-1})(1 - R_{i+d,j+d}) \prod_{k=0}^{d-1} R_{i+k,j+k}, \tag{4}$$

and probability of recurrence of a vertical line of length v is

$$P(d) = \sum_{i,j}^{N} (1 - R_{i,j-1})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k}. \tag{5}$$

The recurrence rate in an RP is

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}. \tag{6}$$

Determinism DET, which is the percentage of recurrence points forming diagonal lines, is

$$DET = \sum_{d=d_{\min}}^{N} dP(d) \left[\sum_{i,j=1}^{N} R_{i,j} \right]^{-1}. \tag{7}$$

Laminarity, which is the percentage of recurrence points forming diagonal lines, is

$$LAM = \sum_{v=v_{\min}}^{N} vP(v) \left[\sum_{i,j=1}^{N} R_{i,j} \right]^{-1}. \tag{8}$$

Entropy $ENTR$, which is the Shannon entropy of the probability distribution of the diagonal line lengths $p(l)$, is defined as

$$ENTR = -\sum_{d=d_{\min}}^{N} p(d) \ln p(d), \text{where } p(d) = \frac{P(d)}{\sum P(d)}. \tag{9}$$

Trapping time TT, which is the average length of the vertical lines, is

$$TT = \sum_{v=v_{\min}}^{N} vp(v) \left[\sum_{v=v_{\min}}^{N} p(v) \right]^{-1} \tag{10}$$

4 Empirical Analysis and Results

Before the VRA and RQA can be performed, the phase space of the potential dynamic system has to be reconstructed from the original time series. Hence in this part we concisely describe the phase space reconstruction technique.

4.1 Phase Space Reconstruction

On a recurrence plot we do not represent the points from a time series, but the distance of states of a dynamic system in a phase space. Therefore, we have to reconstruct the state space from a given time series. As the system of equations generating the trajectory of the system, the original system needs to be reconstructed, which is feasible according to Takens [7]. The most of ten used method for this purpose is the so called time delayed vector method, in which a state vector $y(t)$ can be created from the original time series as follows

$$y(t) = [x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)], \tag{11}$$

where τ is the time delay and m is the embedding dimension, thus τ and m are the embedding parameters. Hence, each point i of the phase space is represented by vector $y(i)$ in a m-dimensional space. In order to properly reconstruct the phase space, two embedding parameters must be appropriately determined.
Time delay. There are two main methods, how to find the optimal time delay. The most often used method is based on the average mutual information (AMI) function which measures the quantity of information attainable from a random variable x for another variable y. AMI $I(\tau)$ is defined as

$$I(\tau) = \sum_{x(t_i), x(t_i+\tau)} p(x(t_i), x(t_i+\tau)) \log_2 \frac{p(x(t_i), x(t_i+\tau))}{p(x(t_i))p(x(t_i+\tau))}, \tag{12}$$

where $p(x(t_i))$, $p(x(t_i+\tau))$ are the probabilities that $x = x(t_i)$ and $y = x(t_i+\tau)$ respectively, and $p(x(t_i), x(t_i+\tau))$ is the probability that $x = x(t_i)$ and $y = x(t_i+\tau)$. The optimal τ is the one that gives minimal $I(\tau)$.
Embedding dimension. There are basically also two methods how to correctly determine embedding dimension. The more preferred alternative for calculating embedding dimension is proposed by Cao [1]. This method is based on statistic

$$E(m) = \frac{1}{N-m\tau} \sum_{i=1}^{N-m\tau} a(t_i, m), \tag{13}$$

where

$$a(t_i, m) = \frac{\| \, y_{m+1}(t_i) - y_{m+1}^{NN}(t_i) \, \|}{\| \, y_m(t_i) - y_m^{NN}(t_i) \, \|}. \tag{14}$$

One can observe that $E(m)$ depends only on m. If m is sufficiently high, $E(m)$ converges to 1, which helps to determine the optimal embedding dimension.

4.2 Empirical results

We use two series of Euro and USD to Czech crown for the analysis. They are daily time series from 1/2010 to 4/2019. The original exchange rate series are transformed into two corresponding series. The first one is the classical logarithmic return series (denoted as Log-returns) and the second one is the series of cumulative daily moments normalized by the maximal value of exchange rate over the whole period (denoted as Momentum series) as follows

$$m_t = \frac{x_t - x_1}{\max(x)}. \tag{15}$$

Their descriptive statistics are shown in Table 1.

In order to reconstruct the phase space of a dynamic system from time series, the embedding parameters must be determined. First, we calculate the optimal time delay τ. To do so, the average mutual information as a function of τ is calculated according to formula (12). The dependency of AMI on time delay for all four series is displayed in Figure 1. From this picture, it is clear that the optimal time delay for log-return series is 1. For the two momentum series the decline of AMI with increasing τ is rather slow. In our opinion, the optimal τ for them can be 6 as the effect of daily momentum may die out after a week.

To determine the optimal embedding dimension, we calculate measure $E2 = E(m+1)/E(m)$, where $E(m)$ is computed according to formula (13). We also calculate an additional measure $E1$ (see [3]) as

$$E1(m) = \frac{E^*(m+1)}{E^*(m)}, \tag{16}$$

	Original series		Log-returns series		Momentum series	
	EURCZK	USDCZK	EURCZK	USDCZK	EURCZK	USDCZK
Mean	26.076	21.259	$-8.90e{-}6$	$9.50e{-}5$	$-3.87e{-}3$	0.119
Median	25.831	20.747	$-3.65e{-}5$	$-6.00e{-}5$	$-1.25e{-}2$	$9.93e{-}2$
Mode	27.021	16.904	0	$-4.01e{-}2$	$2.94e{-}2$	$-4.82e{-}2$
Minimum	23.993	16.235	$-1.66e{-}2$	$-4.01e{-}2$	$-7.73e{-}2$	$-7.39e{-}2$
Maximum	28.374	26.048	$4.62e{-}2$	$5.53e{-}2$	$7.71e{-}2$	0.3028
Std. deviation	1.0336	2.4688	$2.93e{-}3$	$6.61e{-}3$	$3.64e{-}2$	$9.48e{-}2$
Skewness	$-2.31e{-}3$	0.1430	0.1493	0.2124	$-2.31e{-}3$	0.1430
Kurtosis	1.8450	1.9073	31.604	6.6867	1.8450	1.9073
Number of Obs.	2434	2434	2433	2433	2434	2434

Table 1 Descriptive statistics of original and transformed series

where $E^*(m)$ is defined as

$$E^*(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} \mid x(t_i + m\tau) - x^{NN}(t_i + m\tau) \mid . \tag{17}$$

The relationships of $E1, E2$ on m for all four series are shown in Figure 2. From the results displayed in Figure 2 the optimal embedding dimension is 10.

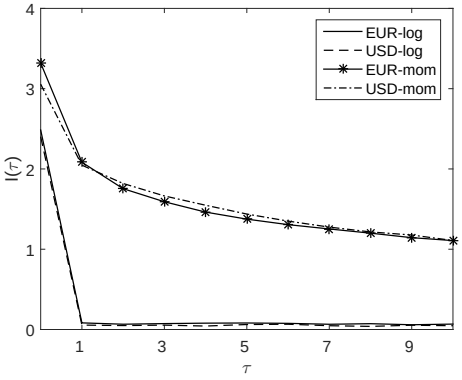
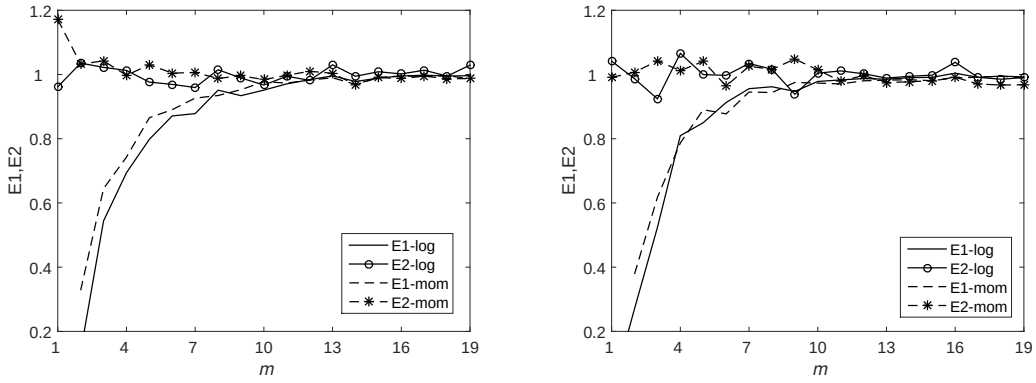


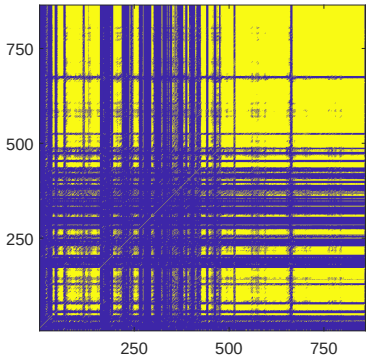
Figure 1 Relationship between average mutual information $I(\tau)$ and time delay τ



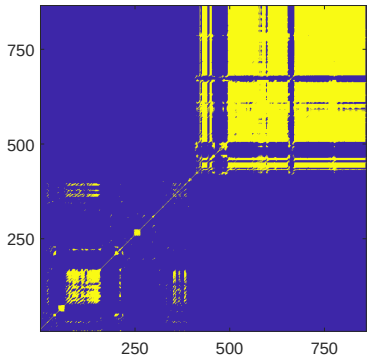
a) EURCZK series b) USDCZK series

Figure 2 Relationship between average mutual information $E1, E2$ and embedding dimension

We use calculated values for embedding parameters to reconstruct the phase space of the system. After the reconstruction, recurrence matrix R is calculated according to formula (1). We do it for do different periods. The

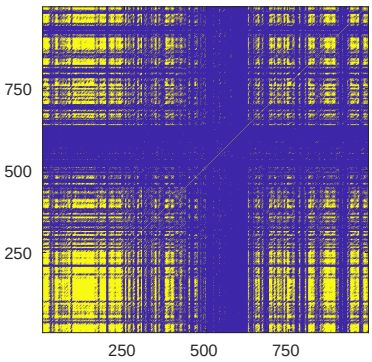


a) Log-return series

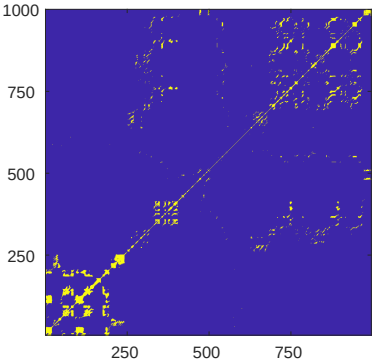


b) Momentum series

Figure 3 Recurrence plot of EURCZK in the period with CNB intervention



a) Log-return series



b) Momentum series

Figure 4 Recurrence plot of EURCZK in period without intervention of the CNB

first on is the period from 2010 to the end of 2013 when the Czech National Bank (CNB) had not intervened to weaken Czech crown (CZK). The second period is from the end of 2013 to 4/2017 in which CNB kept the exchange rate at around 27 CZK per Euro. The recurrence plots of series EURCZK log-returns and momentum[1] are shown in Figures 3 and 4. In these two pictures, the recurrence plot of log-return series are on the left panel and those of momentum series are on the right panel. In the pictures the bright color represent states that reoccurs and the dark color represent the opposites. One can observe the randomness nature of log-return series except the time at the end of intervention period when exchange rate was almost unmoved lightly above level of 27 CZK per Euro. The pictures of period without intervention, the randomness nature is even more apparent with an interval when high volatility is represented by a belt of uninterupted dark color in the left panel. On the right panel, dark color dominates indicating momentum from this period is of greater size than the threshold value. In both cases no periodic patterns are observed excluding the possibility of cyclical fluctuation of exchange rate. Regarding the recurrence plots of log-return and momentum series of daily exchange rate USD to CZK, they exhibit rather characteristic of a random white noise series.

For the four series we analyze, we calculate the quantitative measures of RQA described in section 3 for the four series generated from daily exchange rate in the period with the intervention of the CNB. The results are shown in Table 2. We also calculate these measure for a series of Gaussian white noise with similar statistic characteristics as the logarithmic return series. The results is put in the last column of Table 2.The results in Table 2 show that quantities of RQA of two series derived from daily exchange rate series of USD to CZK do not substantially differ from those of originating from USDCZK exchange rate series. The results of series generated from EURCZK daily exchange rate fully capture the impact of the intervention of the CNB. All quantities of RQA are much higher than those of a Gaussian white noise. In terms of nonlinear time series analysis, the recurrence rate is high. However, this result should be interpreted with caution. The observed recurrence is a result of the intervention which kept

[1] Due to the limited space per each contribution the recurrence plots of USD series cannot be displayed in this paper. However, they can be delivered to those who have interest in them.

	Log-returns series		Momentum series		
Characteristic	EURCZK	USDCZK	EURCZK	USDCZK	White Noise
RR	0.5232	0.0516	0.3523	0.0074	0.0438
DET	0.9611	0.3975	0.9907	0.3899	0.2087
LAM	0.9301	0.2081	1.0321	0.3506	0.4012
ENTR	2.2691	0.5497	3.2366	0.5036	0.3020
TT	19.4703	2.8364	28.3169	2.4790	2.2492

Table 2 Recurrence Quantification Analysis Results

the exchange rate artificially stable rather than it is a consequence of a deterministic process. In our opinion, in order to make more competent inference on character of the dynamics of exchange rate requires more additional research.

5 Conclusion

In order to identify regular patterns in fluctuations of exchange rate, we have chosen the visual recurrence analysis. This technique allows to graphically detect hidden structural changes in data. Then these hidden structures have been quantified by using recurrence quantification analysis (RQA). These two qualitative and quantitative analysis have been applied to exchange rate time series EURCZK and USDCZK from 1/2010 to 4/2019. Data from these period have been converted into two series: logarithmic returns and normalized momentum. We pick two important periods from time series of our interest: period before intervention of CNB and period under under intervention. Using reconstruction technique to obtain the phase space of the system from our time series. The results show that without intervention, our time series seem to behave similarly as a Gaussian white noise. On the other hand, visual analysis can also capture the impact of the intervention. The results of RQA fully confirm the results of qualitative analysis. The results of this research are in agreement with those we did previously for the Czech securities [5]. In order to make full comparison to evaluate ability of nonlinear and linear time series technique, more research is needed.

Acknowledgements

The authors of the paper acknowledge the financial assistance of the University of Economics, Prague as a part of Institutional Research Support no. IP 100040.

References

[1] Cao, L. (1997). Practical method for determining the minimum embedding dimension of scalar time series. *Physica D*, 110(1-2), 43 -50.

[2] Eckmann, J., Kamphorst, S., Ruelle, D. (1987). Recurrence plot of dynamical system. *Europhysics Letters*, 4, 973-977.

[3] Horák, J., Krlín, L.; Raidl, A. (2003). *Deterministický chaos a jeho fyzikální aplikace*. Praha: Academia.

[4] Kennel, M., Brown, R., Abarbanel, H. (1992). Determining embedding dimension for phase-space reconstruction using geometrical construction. *Physical Review A*, 45(6), 3403-3411.

[5] Kodera, J., Tran, Q. V. (2009). Visual recurrence analysis and its application on the Czech stock market. *Politická ekonomie*, 2009(3), 305-322.

[6] Marvan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., Kurths, J. (2002). Recurrence plot based measures of complexity and its application to heart rate variability data. *Physical Review E*, 66(2).

[7] Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*. Berlin: Springer-Verlag. 366–381.

[8] Zbilut, J., Weber, C. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters*, 171, 199-203.

The relation "greater than or equal to" for trapezoidal ordered fuzzy numbers

Krzysztof Piasecki[1], Anna Łyczkowska-Hanćkowiak[2],

Aleksandra Wójcicka-Wójtowicz[3]

Abstract. The ordered fuzzy number (OFN) is defined as a pair of fuzzy number and its orientation. Any OFN is interpreted as imprecise number with additional information about the location of the approximated number. Each positively oriented OFN is interpreted as an imprecise real number described by the linguistic variable "about or slightly above". Each negatively oriented OFN is interpreted as an imprecise real number described by the linguistic variable "about or slightly below". Here, we restrict our considerations to the case of trapezoidal OFN (TrOFN). The main goal of this paper is to introduce the preorder "greater than or equal to" on the space of all TrOFNs. This relation is defined as an extension of analogous relations on the space of all fuzzy numbers. All properties of the introduced relation have been investigated on the basis of the revised Kosiński's theory of OFNs. It is shown here that in the above way this relation has been defined unambiguously as fuzzy ones. In addition, it is proven that the obtained relationship is independent on the orientation of the compared OFNs. The results obtained will be useful for formulating optimization tasks using TrOFNs.

Key words: ordered fuzzy number, fuzzy relation, preorder, strict order.

JEL Classification: C52, C61, C65.
AMS Classification: 03E72, 06A06, 90C70.

1 Introduction

The intuitive concept of ordered fuzzy number (OFN) was introduced by Kosiński and his co-workers [4, 5] as an extension of the concept of fuzzy number (FN) which is interpreted as imprecise approximation of real number. OFNs' usefulness follows from the fact that is interpreted as FN with additional information about the location of the approximated number. For formal reasons, the Kosiński' theory is revised [12] in this way that new definition of OFN fully corresponds to the intuitive defining it by Kosiński. Unfortunately, the OFN theory has one significant drawback. This drawback is the lack of formal mathematical models dedicated to OFN issues.

The purpose of our paper is to fill one of these theoretical gaps. In this paper, we restrict our considerations to the case of trapezoidal OFN (TrOFN). For this reason, here our main aim is to introduce basic theory of ordering TrOFN.

2 Fuzzy number – basic facts

For any space $\mathbb{A}$, the family of all its fuzzy subsets we will denote by the symbol $\mathcal{F}(\mathbb{A})$. An imprecise number is a family of values in which each considered value belongs to it in a varying degree. A commonly accepted model of imprecise number is the FN, defined as a fuzzy subset of the family $\mathbb{R}$ of all real numbers. The most general definition of FN was given by Dubois and Prade [1]. In this paper, we restrict our considerations to the case of TrFNs defined as fuzzy subsets in the space $\mathbb{R}$ of all real numbers in the following way.

Definition 1. *For any nondecreasing sequence* $(a, b, c, d) \subset \mathbb{R}$*, the trapezoidal fuzzy number (TrFN) is the fuzzy subset* $\mathcal{F}(\mathbb{R}) \ni \mathcal{T} = Tr(a, b, c, d)$ *determined explicitly by its membership functions* $\mu_T \in [0,1]^{\mathbb{R}}$ *as follows*

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland
E-mail: krzysztof.piasecki@ue.poznan.pl

[2] WSB University in Poznań, Institute of Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland,
E-mail:anna.lyczkowska-hanckowiak@wsb.poznan.pl

[3] Poznań University of Economics and Business, Department of Operations Research, al. Niepodległości 10, 61-875 Poznań, Poland
E-mail: aleksandra.wojcicka@ue.poznan.pl

$$\mu_T(x) = \mu_{Tr}(x|a,b,c,d) = \begin{cases} 0, & x \notin [a,d], \\ \frac{x-a}{b-a}, & x \in [a,b[, \\ 1, & x \in [b,c], \\ \frac{x-d}{c-d}, & x \in]c,d]. \end{cases} \quad (1)$$

The space of all TrFNs is denoted by the symbol $\mathbb{F}_{Tr}$. The TrFN $Tr(a,a,a,a) = [\![a]\!]$ represents the crisp number $a \in \mathbb{R}$. Therefore, we can write $\mathbb{R} \subset \mathbb{F}_{Tr}$. For any $z \in [b,c]$, the TrFN $Tr(a,b,c,d)$ is interpreted as an imprecise number "about z". Understanding the phrase "about z" depends on the applied pragmatics of the natural language.

Let us take into account any arithmetic operation $*$ defined on $\mathbb{R}$. By $\circledast$ we denote an extension of arithmetic operation $*$ to $\mathbb{F}_{Tr}$. According to the Zadeh's Extension Principle [15], for any pair $(\mathcal{K},\mathcal{L}) \in (\mathbb{F}_{Tr})^2$ represented respectively by their membership functions $\mu_K, \mu_L \in [0,1]^{\mathbb{R}}$, the FN

$$\mathcal{M} = \mathcal{K} \circledast \mathcal{L} \quad (2)$$

is represented by its membership function $\mu_M \in [0,1]^{\mathbb{R}}$ given by the identity:

$$\mu_M(z) = \sup\{\min\{\mu_K(x), \mu_L(y)\} : z = x * y, (x,y) \in \mathbb{R}\}. \quad (3)$$

In line with above, the sum $\oplus$ of TrFNs and difference $\ominus$ between TrFNs are the TrFNs given as follows

$$Tr(a+e, b+f, c+g, d+h) = Tr(a,b,c,d) \oplus Tr(e,f,g,h), \quad (4)$$

$$Tr(a-h, b-g, c-f, d-e) = Tr(a,b,c,d) \ominus Tr(e,f,g,h). \quad (5)$$

In analogous way, the unary minus operator "$-$" on $\mathbb{R}$ is extended to the minus operator $\ominus$ on $\mathbb{F}_{Tr}$ by the identity

$$Tr(-h,-g,-f,-e) = [\![0]\!] \ominus Tr(e,f,g,h) = \ominus Tr(e,f,g,h). \quad (6)$$

Let us consider the pair $(\mathcal{K},\mathcal{L}) \in (\mathbb{F}_{Tr})^2$ of TrFNs represented respectively by their membership functions $\mu_K, \mu_L \in [0,1]^{\mathbb{R}}$. On the set $\mathbb{F}_{Tr}$ of all TrFNs we introduce the relation $\mathcal{K}\ \widetilde{GE}\ \mathcal{L}$, which reads:

"TrFN $\mathcal{K}$ is greater than or equal to TrFN $\mathcal{L}$." (7)

In agreement with the Zadeh's Extension Principle, this relation is a fuzzy order $[GE] \in \mathcal{F}((\mathbb{F}_{Tr})^2)$ determined by its membership function $\nu_{[GE]} \in [0,1]^{(\mathbb{F}_{Tr})^2}$ given as follows [17]

$$\nu_{[GE]}(\mathcal{K},\mathcal{L}) = \sup\{\min\{\mu_K(x), \mu_L(y)\} : x \geq y\}. \quad (8)$$

From the multivalued logic point of view, the value $\nu_{[GE]}(\mathcal{K},\mathcal{L})$ is interpreted as a truth-value of the sentence (8).

Theorem 1. *For any pair $(\mathcal{K},\mathcal{L}) \in (\mathbb{F}_{Tr})^2$ of TrFNs represented respectively by their membership functions $\mu_K, \mu_L \in [0,1]^{\mathbb{R}}$, we have:*

$$\nu_{[GE]}(\mathcal{K},\mathcal{L}) = \nu_{[GE]}(\ominus \mathcal{L}, \ominus \mathcal{K}), \quad (9)$$

$$\nu_{[GE]}(\mathcal{K},\mathcal{L}) = \nu_{[GE]}(\mathcal{K} \ominus \mathcal{L}, [\![0]\!]). \quad (10)$$

Proof. Take into account the quadruple $(\mathcal{K},\mathcal{L},\mathcal{M},\mathcal{N}) \in (\mathbb{F}_{Tr})^4$ of TrFNs represented respectively by their membership functions $\mu_K, \mu_L, \mu_M, \mu_N \in [0,1]^{\mathbb{R}}$.

Let us assume that $\mathcal{M} = \ominus \mathcal{K}$ and $\mathcal{N} = \ominus \mathcal{L}$. Using the identities (3) and (6) we obtain:

$$\mu_M(y) = \mu_K(-y) \quad \text{and} \quad \mu_N(x) = \mu_L(-x).$$

Then the identity (8) implies

$$\nu_{[GE]}(\ominus \mathcal{L}, \ominus \mathcal{K}) = \nu_{[GE]}(\mathcal{N},\mathcal{M}) = \sup\{\min\{\mu_N(x), \mu_M(y)\} : x \geq y\} =$$

$$= \sup\{\min\{\mu_L(-x), \mu_K(-y)\} : -x \leq -y\} = \sup\{\min\{\mu_L(u), \mu_K(v)\} : u \leq v\} = \nu_{[GE]}(\mathcal{K},\mathcal{L}).$$

Let us assume now that $\mathcal{M} = \mathcal{K} \ominus \mathcal{L}$. Using the identity (3) we obtain:

$$\mu_M(z) = \sup\{\min\{\mu_K(x), \mu_L(y)\} : z = x - y, (x,y) \in \mathbb{R}\}.$$

Then the identity (7) implies

$$\nu_{[GE]}(\mathcal{K} \ominus \mathcal{L}, [\![0]\!]) = \nu_{[GE]}(\mathcal{M}, [\![0]\!]) = \sup\{\mu_M(z) : z \geq 0\} =$$

$$= \sup\{\sup\{\min\{\mu_K(x), \mu_L(y)\} : z = x - y, (x,y) \in \mathbb{R}\} : z \geq 0\} =$$

$$= \sup\{\min\{\mu_K(x), \mu_L(y)\} : x - y \geq 0\} = \nu_{[GE]}(\mathcal{K},\mathcal{L}). \quad \text{QED}$$

Theorem 2. *For any TrFNs* $Tr(a,b,c,d), Tr(e,f,g,h) \in \mathbb{F}_{Tr}$ *we have*

$$\nu_{[GE]}\big(Tr(a,b,c,d), Tr(e,f,g,h)\big) =$$

$$= \nu_{[GE]}(Tr(a-h,b-g,c-f,d-e),[\![0]\!]) = \varphi(c,d,e,f) = \begin{cases} 0, & 0 > a-h, \\ \frac{h-a}{b+h-g-a}, & a-h \geq 0 > b-g, \\ 1, & b-g \geq 0. \end{cases} \quad (11)$$

Proof. Successively from (10) and (8), we get

$$\nu_{[GE]}\big(Tr(a,b,c,d), Tr(e,f,g,h)\big) = \nu_{[GE]}(Tr(a-h,b-g,c-f,d-e),[\![0]\!]) = \varphi(a,b,g,h) \ . \text{QED}$$

3 Ordered fuzzy numbers

In this paper, we restrict our considerations to the case of TrOFN defined in the following way.

Definition 2. [12] *For any monotonic sequence* $(a,b,c,d) \subset \mathbb{R}$, *the trapezoidal ordered fuzzy number (TrOFN)* $\overrightarrow{Tr}(a,b,c,d) = \overrightarrow{\mathcal{T}}$ *is the pair of the orientation* $\overrightarrow{a,d} = (a,d)$ *and fuzzy subset* $\mathcal{T} \in \mathcal{F}(\mathbb{R})$ *determined explicitly by its membership functions* $\mu_T \in [0,1]^{\mathbb{R}}$ *as follows*

$$\mu_T(x) = \mu_{Tr}(x|a,b,c,d) = \begin{cases} 0, & x \notin [a,d] \equiv [d,a], \\ \frac{x-a}{b-a}, & x \in [a,b[\equiv]b,a], \\ 1, & x \in [b,c] \equiv [c,b], \\ \frac{x-d}{c-d}, & x \in]c,d] \equiv [d,c[. \end{cases} \quad (12)$$

The space of all TrOFNs is denoted by the symbol $\mathbb{K}_{Tr}$. Any TrOFN is interpreted as imprecise number with additional information about the location of the approximated number. This information is given as orientation of TrOFN. The fulfilment of the condition $a < d$ determines the positive orientation of TrOFN $\overrightarrow{Tr}(a,b,c,d)$. For any $z \in [b,c]$, the positively oriented TrOFN $Tr(a,b,c,d)$ is interpreted as an imprecise number "about or slightly above z". The space of all positively oriented TrOFN is denoted by the symbol $\mathbb{K}_{Tr}^{+}$. The fulfilment of the condition $a > d$ determines the negative orientation of TrOFN $\overrightarrow{Tr}(a,b,c,d)$. For any $z \in [c,b]$, the negatively oriented TrOFN $Tr(a,b,c,d)$ is interpreted as an imprecise number "about or slightly below z". The space of all negatively oriented OFN is denoted by the symbol $\mathbb{K}_{Tr}^{-}$. For the case $a = d$, TrOFN $\overrightarrow{Tr}(a,a,a,a) = [\![a]\!]$ represents a crisp number $a \in \mathbb{R}$, which is not oriented. Understanding the phrases "about or slightly above z" and "about or slightly below z" depend on the applied pragmatics of the natural language. Summing up, we can write

$$\mathbb{K}_{Tr} = \mathbb{K}_{Tr}^{+} \cup \mathbb{R} \cup \mathbb{K}_{Tr}^{-}. \quad (13)$$

The difference $\boxminus$ between TrOFNs is the TrOFN given as follows

$$\overrightarrow{Tr}(a,b,c,d) \boxminus \overrightarrow{Tr}(a-p,b-q,c-r,d-s) =$$

$$= \begin{cases} \overrightarrow{Tr}(\min\{p,q\},q,r,\max\{r,s\}) & (q < r) \vee (q = r \wedge p \leq s) \\ \overrightarrow{Tr}(\max\{p,q\},q,r,\min\{r,s\}) & (q > r) \vee (q = r \wedge p > s). \end{cases} \quad (14)$$

The unary minus operator "$-$" on $\mathbb{R}$ is extended to the minus operator $\boxminus$ on $\mathbb{K}_{Tr}$ by the identity

$$\overrightarrow{Tr}(-a,-b,-c,-d) = [\![0]\!] \boxminus \overrightarrow{Tr}(a,b,c,d) = \boxminus \overrightarrow{Tr}(a,b,c,d). \quad (15)$$

For the case $a \geq d$ the membership function of TrFN $Tr(a,b,c,d)$ is equal to the membership function of TrOFN $\overrightarrow{Tr}(a,b,c,d)$. This fact implies the existence of isomorphism $\Psi\colon (\mathbb{K}_{Tr}^{+} \cup \mathbb{R}) \to \mathbb{F}_{Tr}$ given by the identity

$$Tr(a,b,c,d) = \Psi\left(\overrightarrow{Tr}(a,b,c,d)\right). \quad (16)$$

This isomorphism may be extended to the space $\mathbb{K}_{Tr}$ by disorientation map $\overline{\Psi}\colon \mathbb{K}_{Tr} \to \mathbb{F}_{Tr}$ given by the identity

$$\overline{\Psi}(\overrightarrow{\mathcal{K}}) = \begin{cases} \Psi(\overrightarrow{\mathcal{K}}) & \overrightarrow{\mathcal{K}} \in \mathbb{K}_{Tr}^{+} \cup \mathbb{R}, \\ \ominus \Psi(\boxminus \overrightarrow{\mathcal{K}}) & \overrightarrow{\mathcal{K}} \in \mathbb{K}_{Tr}^{-}. \end{cases} \quad (17)$$

Lemma 1. *For any pair* $(\overrightarrow{\mathcal{K}}, \overrightarrow{\mathcal{L}}) \in (\mathbb{K}_{Tr}^{+} \cup \mathbb{R}) \times \mathbb{K}_{Tr}^{-}$ *we have*

$$\Psi(\overrightarrow{\mathcal{K}} \boxminus \overrightarrow{\mathcal{L}}) = \Psi(\overrightarrow{\mathcal{K}}) \ominus \left(\ominus \Psi(\boxminus \overrightarrow{\mathcal{L}})\right). \quad (18)$$

Proof. Let $\overrightarrow{\mathcal{K}} = \overrightarrow{Tr}(a,b,c,d) \in (\mathbb{K}_{Tr}^{+} \cup \mathbb{R})$ and $\overleftarrow{\mathcal{L}} = \overleftarrow{Tr}(e,f,g,h) \in \mathbb{K}_{Tr}^{-}$. Then we have $\boxminus \overleftarrow{\mathcal{L}}, \overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}} \in \mathbb{K}_{Tr}^{+}$ because of the sequences $(-e,-f,-g,-h)$ and $(a-e, b-f, c-g, d-h)$ are nondecreasing. Then, from (14) we get

$$\Psi\left(\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}}\right) = \Psi\left(\overrightarrow{Tr}(a-e, b-f, c-g, d-h)\right) = Tr(a-e, b-f, c-g, d-h).$$

On the other hand, successively from (17), (15), (17), (6) and (5), we obtain

$$\Psi\left(\overrightarrow{\mathcal{K}}\right) \ominus \left(\ominus \Psi\left(\boxminus \overleftarrow{\mathcal{L}}\right)\right) = \Psi\left(\overrightarrow{Tr}(a,b,c,d)\right) \ominus \left(\ominus \Psi\left(\boxminus \overrightarrow{Tr}(e,f,g,h)\right)\right) =$$

$$= Tr(a,b,c,d) \ominus \left(\ominus \Psi\left(\overrightarrow{Tr}(-e,-f,-g,-h)\right)\right) = Tr(a,b,c,d) \ominus \left(\ominus Tr(-e,-f,-g,-h)\right) =$$

$$= Tr(a,b,c,d) \ominus Tr(h,g,f,e) = Tr(a-e, b-f, c-g, d-h). \text{ QED}$$

4 Ordering of order fuzzy numbers

Let us consider the pair $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr})^2$ represented respectively by the pair $(\mu_K, \mu_L) \in ([0,1]^{\mathbb{R}})^2$ of their membership functions. On the set $\mathbb{K}_{tr}$ of all OFNs we introduce the relation $\overrightarrow{\mathcal{K}}\ \widetilde{GE}\ \overleftarrow{\mathcal{L}}$, which reads:

$$"TrOFN\ \overrightarrow{\mathcal{K}}\ is\ greater\ than\ or\ equal\ to\ TrOFN\ \overleftarrow{\mathcal{L}}." \tag{19}$$

This relation is a fuzzy order $\widetilde{GE} \in \mathcal{F}((\mathbb{K}_{Tr})^2)$ defined by means of its membership function $\nu_{GE} \in [0,1]^{(\mathbb{K}_{Tr})^2}$ fulfilling the conditions:

- for any pair $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^{+} \cup \mathbb{R})^2$, the extension principle

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{[GE]}\left(\Psi\left(\overrightarrow{\mathcal{K}}\right), \Psi\left(\overleftarrow{\mathcal{L}}\right)\right), \tag{20}$$

- for any pair $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr})^2$, the sign exchange law

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{GE}\left(\boxminus \overleftarrow{\mathcal{L}}, \boxminus \overrightarrow{\mathcal{K}}\right), \tag{21}$$

- for any pair $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr})^2$, the law of parties' subtraction of inequality

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{GE}\left(\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}}, [\![0]\!]\right). \tag{22}$$

Theorem 3. *For any pair* $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr})^2$ *we have*

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{[GE]}\left(\overline{\Psi}\left(\overrightarrow{\mathcal{K}}\right), \overline{\Psi}\left(\overleftarrow{\mathcal{L}}\right)\right). \tag{23}$$

Proof. For any pair $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^{+} \cup \mathbb{R})^2$ the identity (23) is obvious. Let us assume that $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^{-})^2$. Then, $\left(\boxminus \overrightarrow{\mathcal{K}}, \boxminus \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^{+})^2$ and successively from (21), (20), (9) and (17), we get

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{GE}\left(\boxminus \overleftarrow{\mathcal{L}}, \boxminus \overrightarrow{\mathcal{K}}\right) = \nu_{[GE]}\left(\Psi\left(\boxminus \overleftarrow{\mathcal{L}}\right), \Psi\left(\boxminus \overrightarrow{\mathcal{K}}\right)\right) = \nu_{[GE]}\left(\ominus \Psi\left(\boxminus \overrightarrow{\mathcal{K}}\right), \ominus \Psi\left(\boxminus \overleftarrow{\mathcal{L}}\right)\right) =$$

$$= \nu_{[GE]}\left(\overline{\Psi}\left(\overrightarrow{\mathcal{K}}\right), \overline{\Psi}\left(\overleftarrow{\mathcal{L}}\right)\right).$$

Let us assume now that $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in (\mathbb{K}_{Tr}^{+} \cup \mathbb{R}) \times \mathbb{K}_{Tr}^{-}$. Then $\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}} \in \mathbb{K}_{Tr}^{+}$ and successively from (22), (20), (18), (22) and (17), we get

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{GE}\left(\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}}, [\![0]\!]\right) = \nu_{[GE]}\left(\Psi\left(\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}}\right), [\![0]\!]\right) = \nu_{[GE]}\left(\Psi\left(\overrightarrow{\mathcal{K}}\right) \ominus \left(\ominus \Psi\left(\boxminus \overleftarrow{\mathcal{L}}\right)\right), [\![0]\!]\right) =$$

$$= \nu_{[GE]}\left(\Psi\left(\overrightarrow{\mathcal{K}}\right), \ominus \Psi\left(\boxminus \overleftarrow{\mathcal{L}}\right)\right) = \nu_{[GE]}\left(\overline{\Psi}\left(\overrightarrow{\mathcal{K}}\right), \overline{\Psi}\left(\overleftarrow{\mathcal{L}}\right)\right).$$

Let us assume now that $\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) \in \mathbb{K}_{Tr}^{-} \times (\mathbb{K}_{Tr}^{+} \cup \mathbb{R})$. Then $\overleftarrow{\mathcal{L}} \boxminus \overrightarrow{\mathcal{K}} \in \mathbb{K}_{Tr}^{+}$ and successively from (22), (21), (20), (18), and (17), we get

$$\nu_{GE}\left(\overrightarrow{\mathcal{K}}, \overleftarrow{\mathcal{L}}\right) = \nu_{GE}\left(\overrightarrow{\mathcal{K}} \boxminus \overleftarrow{\mathcal{L}}, [\![0]\!]\right) = \nu_{GE}\left([\![0]\!], \overleftarrow{\mathcal{L}} \boxminus \overrightarrow{\mathcal{K}}\right) = \nu_{[GE]}\left([\![0]\!], \Psi\left(\overleftarrow{\mathcal{L}} \boxminus \overrightarrow{\mathcal{K}}\right)\right) =$$

$= \nu_{[GE]}\left([\![0]\!], \Psi(\overrightarrow{\mathcal{L}}) \ominus \left(\ominus \Psi(\boxminus \overrightarrow{\mathcal{K}})\right)\right) = \nu_{[GE]}\left(\ominus \Psi(\boxminus \overrightarrow{\mathcal{K}}), \Psi(\overrightarrow{\mathcal{L}})\right) = \nu_{[GE]}\left(\overline{\Psi}(\overrightarrow{\mathcal{K}}), \overline{\Psi}(\overrightarrow{\mathcal{L}})\right)$. QED

For any finite set $A = \{\overrightarrow{\mathcal{K}}_1, \overrightarrow{\mathcal{K}}_2, \dots, \overrightarrow{\mathcal{K}}_n\} \subset \mathbb{K}_{Tr}$ we can distinguish its set of maximal elements given as fuzzy subset $\mathrm{Max}\{A\} \in \mathcal{F}(A)$. The set $\mathrm{Max}\{A\}$ of maximal element is determined by its membership function $\mu_{\mathrm{Max}\{A\}} \in [0,1]^A$ given by the identity (Orlovsky, 1979)

$$\mu_{\mathrm{Max}\{A\}}(\overrightarrow{\mathcal{K}}_i) = \min\{\nu_Q(\overrightarrow{\mathcal{K}}_i, \overrightarrow{\mathcal{K}}_j) : \overrightarrow{\mathcal{K}}_j \in A\}. \tag{24}$$

This set may be applied as solution of optimization tasks using TrOFNs. Moreover, let us note, that the set $\mathrm{Max}\{A\}$ of maximal elements may be applied as a fuzzy choice function [2].

5 Numerical example

Let us take into account finite set $A = \{\overrightarrow{\mathcal{K}}_1, \overrightarrow{\mathcal{K}}_2, \overrightarrow{\mathcal{K}}_3, \overrightarrow{\mathcal{K}}_4\} \subset$, where

$\overrightarrow{\mathcal{K}}_1 = \overrightarrow{Tr}(10,20,30,40)$, $\overrightarrow{\mathcal{K}}_2 = \overrightarrow{Tr}(5,25,40,60)$, $\overrightarrow{\mathcal{K}}_3 = \overrightarrow{Tr}(15,15,10,5)$, $\overrightarrow{\mathcal{K}}_4 = \overrightarrow{Tr}(25,15,10,5)$.

In the first step, we transform each OFN using the disorientation map (17). We get

$\overline{\Psi}(\overrightarrow{\mathcal{K}}_1) = \overrightarrow{Tr}(10,20,30,40), \overline{\Psi}(\overrightarrow{\mathcal{K}}_2) = \overrightarrow{Tr}(5,25,40,60), \overline{\Psi}(\overrightarrow{\mathcal{K}}_3) = \overrightarrow{Tr}(5,10,15,15), \overline{\Psi}(\overrightarrow{\mathcal{K}}_4) = \overrightarrow{Tr}(5,10,15,25)$.

In the next, for any pair $(\overrightarrow{\mathcal{K}}_i, \overrightarrow{\mathcal{L}}_j) = (\overrightarrow{\mathcal{K}}_i, \overrightarrow{\mathcal{K}}_j) \in A^2$, we calculate the difference $\mathcal{M}_{i,j} = \overline{\Psi}(\overrightarrow{\mathcal{K}}_i) \boxminus \overline{\Psi}(\overrightarrow{\mathcal{L}}_j)$ with the use (5). The results of these calculations are presented in Table 1.

	$\overrightarrow{\mathcal{K}}_1$	$\overrightarrow{\mathcal{K}}_2$	$\overrightarrow{\mathcal{K}}_3$	$\overrightarrow{\mathcal{K}}_4$
$\overrightarrow{\mathcal{L}}_1$	$Tr(-30,-10,10,30)$	$Tr(-30,-5,20,50)$	$Tr(-35,-20,-5,5)$	$Tr(-35,-20,-5,15)$
$\overrightarrow{\mathcal{L}}_2$	$Tr(-50,-20,5,30)$	$Tr(-55,-15,15,55)$	$Tr(-10,10,30,55)$	$Tr(-20,10,30,55)$
$\overrightarrow{\mathcal{L}}_3$	$Tr(-5,5,20,35)$	$Tr(-10,10,30,55)$	$Tr(-10,-5,5,10)$	$Tr(-10,-5,5,20)$
$\overrightarrow{\mathcal{L}}_4$	$Tr(-15,5,20,35)$	$Tr(-20,10,30,55)$	$Tr(-20,-5,5,10)$	$Tr(-20,-5,5,20)$

Table 1 Differences $\mathcal{M}_{i,j}$ between compared TrOFNs

In the end, using identity (11), for any pair $(\overrightarrow{\mathcal{K}}_i, \overrightarrow{\mathcal{L}}_j) \in A^2$ of TrOFNs, we calculate the membership function value $\nu_{[GE]}\left(\overline{\Psi}(\overrightarrow{\mathcal{K}}_i), \overline{\Psi}(\overrightarrow{\mathcal{L}}_j)\right)$ evaluating the sentence (9). The results of these calculations are presented in Table 2. The set $\mathrm{Max}\{A\}$ of maximal elements is determined by its membership function $\mu_{\mathrm{Max}\{A\}} \in [0,1]^A$ given by the identity (42). The values $\mu_{\mathrm{Max}\{A\}}(\overrightarrow{\mathcal{K}}_i)$ are shown in the last row of Table 2.

	$\overrightarrow{\mathcal{K}}_1$	$\overrightarrow{\mathcal{K}}_2$	$\overrightarrow{\mathcal{K}}_3$	$\overrightarrow{\mathcal{K}}_4$
$\overrightarrow{\mathcal{L}}_1$	1	1	0.5	0.75
$\overrightarrow{\mathcal{L}}_2$	1	1	1	1
$\overrightarrow{\mathcal{L}}_3$	1	1	1	1
$\overrightarrow{\mathcal{L}}_4$	1	1	1	1
$\mu_{\mathrm{Max}\{A\}}(\overrightarrow{\mathcal{K}}_i)$	1	1	0.5	0.75

Table 2 The membership functions $\nu_{[GE]}\left(\overline{\Psi}(\overrightarrow{\mathcal{K}}_i), \overline{\Psi}(\overrightarrow{\mathcal{L}}_j)\right)$ and $\mu_{Max\{A\}}(\overrightarrow{\mathcal{K}}_i)$

6 Final remarks

Relation "greater than or equal to" $\widetilde{GE}$ is explicitly defined on the space of all TrOFNs. We proved that the relation $\widetilde{GE}$ is independent of the orientation of the numbers being compared. This conclusion may be applied for simplification of many TrOFN applications. The OFNs' ordering obtained here will allow us to use the financial

ratios proposed in the investment decision making models described in [6,7,8, 11]. Moreover, this ordering may be applied for decision making problems described in [3, 10, 13, 14].

All results presented here should be generalized to any OFN case. The only difficulty here is to prove Lemat 1.

References

[1] Dubois D., Prade H. (1979). Fuzzy real algebra: some results, *Fuzzy Sets and Systems*, 2, 327-348, https://doi.org/10.1016/0165-0114(79)90005-8 .

[2] Herrera F., Herrera-Viedma E. (2000). Choice functions and mechanisms for linguistic preference relations, *European Journal of Operational Research*, 120(1), 144-161, https://doi.org/10.1016/S0377-2217(98)00383-X .

[3] Kacprzak D., Kosiński W., Kosiński W.K. (2013) . Financial Stock Data and Ordered Fuzzy Numbers, In L. Rutkowski et al. (Eds.), *12th International Conference, ICAISC 2013, Proceedings, Part I,* Lecture Notes in Artificial Intelligence 7894. (pp. 259–270). Berlin Heidelberg: Springer-Verlag.

[4] Kosiński, W., Prokopowicz, P., Ślęzak D. (2002). Fuzzy numbers with algebraic operations: algorithmic approach. In: M. Kłopotek et al., (Eds.) *Proc.IIS'2002,* (pp. 311-320), Heidelberg: Physica Verlag.

[5] Kosiński, W. (2006). *On fuzzy number calculus*. *Int. J. Appl. Math. Comput. Sci.*, 16(1), 51–57.

[6] Łyczkowska-Hanćkowiak A. (2019). Sharpe's Ratio for Oriented Fuzzy Discount Factor, *Mathematics*, 7(3), 272 , https://doi.org/10.3390/math7030272

[7] Łyczkowska-Hanćkowiak A. (2019). Jensen's Ratio for Oriented Fuzzy Discount Factor, In *37th International Conference Mathematical Methods in Economics. Conference Proceedings* (under review)

[8] Łyczkowska-Hanćkowiak A., Piasecki K. (2019) . The Treynor's ratio for discount factor given as ordered trapezoidal fuzzy number, In W. Szkutnik et al (Eds.), *11th International Scientific Conference "Analysis of International Relations 2019. Methods and Models of Regional Development. Winter Edition". Conference Proceedings*, (pp.31-44), Katowice: Publishing House of the University of Economics in Katowice.

[9] Orlovsky S.A. (1978). Decision making with a fuzzy preference relation, *Fuzzy Sets and Systems*, 1(3), 155-167, https://doi.org/10.1016/0165-0114(78)90001-5 .

[10] Prokopowicz P. et al (Eds.) (2017). *Theory and Applications of Ordered Fuzzy Number. Tribute to Professor Witold Kosiński*, Studies in Fuzziness and Soft Computing 356, Springer Verlag, Berlin.

[11] Piasecki, K. (2017). Expected return rate determined as oriented fuzzy number, In P. Prazak (Ed.) *35th International Conference Mathematical Methods in Economics. Conference Proceedings*, (pp. 561-565), Hradec Kralove, Gaudeamus.

[12] Piasecki K., (2018). Revision of the Kosiński's theory of ordered fuzzy numbers, *Axioms* 7(1), 16, https://doi.org/10.3390/axioms7010016 .

[13] Piasecki K, Roszkowska E. (2018). On Application of Ordered Fuzzy Numbers in Ranking Linguistically Evaluated Negotiation Offers, *Advances in Fuzzy Systems*, 1, 1-12, https://doi.org/10.1155/2018/1569860 .

[14] Roszkowska E., Kacprzak D. (2016). *The fuzzy SAW and fuzzy TOPSIS procedures based on ordered fuzzy numbers,* "Information Sciences", Vol. 369, pp 564-584 https://doi.org/10.1016/j.ins.2016.07.044 .

[15] Zadeh L. (1965). *Fuzzy sets, Information and Control*, 8, pp 338-353, https://doi.org/10.1016/s0019-9958(65)90241-x .

Adaptive population techniques in Evolution Algorithms

František Koblasa[1], Miroslav Vavroušek[2]

Abstract. Evolution algorithms are one of the most popular optimization techniques for its ability to solve wide field of general combinatorial problems. They are simple to implement and use as its mechanisms are logical and untestable due to its natural and common sense. However its optimization results as well as timespan are heavy dependent on setting searching parameters as crossover, mutation rates, population size etc. Setting those parameters requires intuition and experience and is also beneficial to change them during optimization as necessity to exploit and explore changes during optimization. Those are the reasons why nowadays research in the field of Evolutionary Computing is focusing more and more on the adaptive operator control.

This article is focusing on part of adaptive parameter control which is known not only as responsible for quality of results but also for optimization time. The research is focusing on population sizing schemes together with selection and elimination procedures. It is reviewing known techniques and presenting original solution. Those techniques are then tested on known theoretical job shop scheduling problems and their efficiency is discussed.

Keywords: Evolution algorithms, Selection, Population size, Elimination, Job shop scheduling.

JEL Classification: C63
AMS Classification:90C59

1 Introduction

Evolution algorithms (EA) are considered as most versatile optimization algorithms which doesn't need specific oriented knowledge (beside memetic and hybrid EA) in comparison to local search algorithm as taboo search or other knowledge based branch and bound schemes. Performance of EAs is usually given by numerous operators setting [17]. That setting is usually constant during optimization and dependent on algorithm and size of the problem. This is known as parameter tuning problem [6]. This problem is addressed very frequently across EAs applications not just in parameter volume but also in scope of operator selection [21]. However, the optimal parameter values are changing during optimization process. This is known as the parameter control problem. The benefits of online (changing during optimization process) parameter settings are known for some time [4]. It allows EA to change searching strategy during optimization as at beginning it is beneficial to search in more wide approach and in the end in depth. The algorithm can to adjust to the changing fitness landscapes when solving dynamic problems. It allows algorithm to analyze genotype or phenotype information and adjust searching strategy according to e.g. stuck in local optimum or convergence of whole population to one solution. The most important benefit from the point of view of EAs applicability in practice is that algorithm sets setting itself, not the user. Those benefits were main motivation to test well known parameter adaptation techniques with classical static and to compare them with newly proposed one. This research is dedicated to use those techniques in the production planning and scheduling. That is why paper is organized as follows.

The first chapter defines Job Shop Scheduling Problems and general approach to solve it. JSSP is used as model to test EA optimization techniques. These techniques, including newly proposed, are than defined in following chapter. Experiment chapter is than focusing on ability to obtain best possible solution in given time as biggest motivation of this research is applicability in practice. Future research direction in adaptable evolution based algorithms is that discussed.

Technical university of Liberec, Department of manufacturing systems and automation, Studentská 2, Liberec 1, Czech Republic,
[1] frantisek.koblasa@tul.cz
[2] miroslav.vavrousek@tul.cz

2 Job shop Scheduling problem

Job shop Scheduling is the combinatorial and in the most cases NP-Hard problem (JSSP), in which multiple jobs n are processed on several machines m. Each job$\{J_i\}_{1\leq j\leq n}$consists of a sequence of tasks O_{jr} with processing time P_{jr}, which must be performed in a given order, and each task must be processed on a specific machine $\{M_r\}_{1\leq r\leq m}$. The classical scheduling $n\times m$ problems are usually minimizing objective function of makespan (1), e.g. time required to complete all the jobs, where completion times for each operation $\{c_{jr}\}_{1\leq j\leq n,\,1\leq r\leq m}$

$$C_{max} = \max(c_{1,1}, \dots, c_{n,m}) \qquad (1)$$

Solution together with fitness function is calculated by constructive Giffler and Thompson active schedule generation algorithm [12].

There are unlimited ways how to approach on this sequencing decision process beginning with the classical and in the practice most used dispatching (priority) rules continuing with searching techniques as popular Taboo method ending with Evolution Algorithms and Artificial Intelligence. Further chapter is focusing on using Evolution algorithms and specifically setting necessary parameters.

3 Evolution Algorithms population techniques

Evolutionary Algorithms are meta-heuristic techniques used to solve problems that cannot be easily solved in polynomial time e.g. NP-Hard problems as the one described in the previous chapter, or polynomial problems with large searching space. Evolution Algorithm or it´s subclass Genetic Algorithm (GA) has typically several stages which are called often operators:

1. Initialization – generation of usually random generated [9] chromosomes in the form of vector, representing solution to given problem. There are number of possible representations which lots of them are problem specific. The type of solution representation often limits other operators like crossover type or the mutation [18]. After solutions are generated, it is necessary to apply constructive algorithm to evaluate fitness function. There is used random key based representation in all tested GA. During initialization there is generated initially defined number of individuals μ. Non-adaptive algorithms have fixed μ during whole optimization. Adaptive population sizing changes size of the population base on dynamic elimination strategy.

2. Selection – procedure to select solutions (parents) which will be recombined in to the new ones (offspring – childes). There are usually selected λ-times 2 parents generating 2 childes (2+2 mating), however there are techniques where more parents are used [5] to create one individual. There are several possible ways to select them as completely randomly [5], selection based on fitness quality or genotype [11, 22]. Further presented GA are using different approaches, however selection is not changing dynamically (adaptively). Number of selected parents is given by evolution strategy. There are simple strategies where only mutation is applied [8] on one individual, most of them use some form of multiple selection of λ pairs.

3. Reproduction – crossover and mutation are the most used mechanisms of reproduction. Both of them require at least two operator setups. The first is type of operator and second is probability of its occurrence. There are t_c types of crossover beginning with point crossover, splitting chromosome in parts, to uniform where crossing operation is done from gene to gene. The second - t_m mutation than can be divided to random change of gene and to knowledge based (memetic algorithms) system which uses problem specific manipulation [10]. Nevertheless, crossover as well as mutation requires setting probability that operation would be done on gene (chromosome) and are usually defined as p_c (crossover) and p_m (mutation) as threshold which passing activates crossover (p_c = 0.85 means that there is 15% probability of gene crossing).

4. Elimination – procedure to decide which individuals will survive in to next generation. The most used principles include some form of elitist strategy, which means survivability depends on quality of fitness function. The most used are selecting better half of $\mu+\lambda$ individuals. There is hardly ever some strategy which does not save permanently the best solution [20]. Elimination together with selection is main core of evolution strategy and it is described in detail on examples of each applied algorithms.

5. Repetition – algorithm repeats optimization process beginning from 2nd step of selection for G number of generation.

Each algorithm can differ from described scheme. Differences together with additional operators are described in following subchapters.

3.1 Simple genetic and biased genetic algorithm

Simple genetic algorithm (SGA) follows previously described pattern. Random key representation of the JSSP problem is used [3]. Uniform crossover is used to generate new individual. This research does not follow pattern of knowledge based mutation (e.g. Critical path/block search) so diversity of population should be kept by randomness in representation. There are selected two kinds of parent selection, classical – totally random for SGA and biased roulette wheel for biased genetic algorithm (BSGA). Biased [13] selection uses probability to be picked base on quality of the fitness function. The first pick select parent from the best solutions (1), the second one from the worst (2).

$$P(x_i) = \frac{\sum_{j=1}^{N} f_j - f(x_i)}{\sum_{j=1}^{N} f_j} \qquad (1) \qquad \text{and } P(x_i) = \frac{f(x_i)}{\sum_{j=1}^{N} f_j} \qquad (2)$$

Elitist elimination strategy is then used, so better half survives according objective function. Both of those algorithms (SGA and BSGA) represent non adaptive algorithm and suites as comparison to highlight benefits and disadvantages of adaptive algorithms below. Every other algorithm is following SGA if not said otherwise.

3.2 Aging individual algorithms

Genetic Algorithm with Varying Population Size (GAVaPS) [1] and Adaptation Population size Genetic Algotithm (APGA) [2] represents algorithms where elimination process is done by mechanism of aging.

GAVaPS base on the fitness function sets for every new individual *i* lifetime according to (3) when $fitness(i) \geq AvgFit$ else (4), where *MinLT* and *MaxLT* represents minimal and maximal allowable age.

$$lifetime(i) = MinLT + \frac{1}{2}(MaxLT - MinLT)\frac{WortsFitness - fitness(i)}{WortsFitness - AvgFit} \qquad (3)$$

$$lifetime(i) = \frac{1}{2}(MinLT + MaxLT) + \frac{1}{2}(MaxLT - MinLT)\frac{AvgFit - fitness(i)}{AvgFit - MaxFit} \qquad (4)$$

The main difference between GAVaPS and APGA is that in APGA the best individual does not age and in GAVaPS the best individual can die of age (but is saved as best). Also APGA has usually higher *MaxLT*. Both of them use point crossover, GAVaPS one point and APGA two point system. GAVaPS is using random selection, however APGA tournament. APGA comparing to other GAs is using additional probability that selected parents will agree to mate. If not new individuals are made by mutation, not crossover. However, as this research is not using mutation operator, every selected pair of parent will mate. On the contrary APGA is using strategy where each generation, only two parents are selected to reproduce. inGAVaPS which eliminates clones base on fitness function is selected as representative of aging algorithms to be tested.

3.3 Self-adaptive algorithm

Self-adaptive algorithm (SAGA) [15] basic idea is to run three independent SGAs, each with its own population (let´s call them P1, P2, P3 with size (P1) < size (P2) < size (P3)).There are rules that adjust the population sizes in order to "optimize" the performance of the mid-sized population P2. There are multiple rules to optimize size of P2 populations if:

- P1&P3 have the same fitness $size(P1) = 0.5size(P1)$ and $size(P3) = 2size(P3)$
- P1&P2 have the same fitness: $size(P1) = 0.5size(P1)$
- P2&P3 have the same fitness: $size(P3) = 2size(P3)$
- P1>P2>P3 in the terms of fitness quality :
  $size(P1) = size(P2), size(P2) = size(P3), size(P3) = 2size(P3)$
- P1>P3>P2 or P2>P1>P3 : $size(P1) = 0.5(size(P1) + size(P2))$
- P2>P3 P1 or P3>P1>P2 : $size(P3) = 0.5size(P2) + size(P3)$
- P3>P2>P1: $size(P1) = 0.5size(P1), size(P2) = size(P1), size(P3) = size(P2)$

3.4 Population Resizing on Fitness Improvement

Population Resizing on Fitness Improvement GA (PRoFIGA) in this research represent adaptation base on improvement speed of the population. PRoFIGA resize the population by growing or shrinking based on an improvement of the best fitness contained in the population. Population grows when there is an improvement of the best fitness function or when there is no improvement in defined number of iteration (in our case 5) else population shrinks (decrease factor = increase factor). Growing factor X is than (5): *IF* –increase factor is usually between (0,1), *MG* –maximal number of generation, which in our case is initially set to 1000 and actualized every generation. That is based on forecast of previous population number, its timespan calculation and overall remaining timespan. *CG* is number of current generation, max*FIT(i)* is current best fitness, max*FITp* –previous best fitness, *inibestFIT* –initial best fitness.

$$X = IF * (MG - CG) * \frac{BestFITp - bestFIT(i)}{inibestFIT} \qquad (5)$$

3.5 APSC

Adaptive Population, Selection and Crossover (APSC) GA is using knowledge from previously reviewed algorithms and it is adding knowledge oriented crossover. APSC is using population sizing scheme from PRoFIGA. Several parents from this population are selected by biased scheme to produce new offspring by crossover. Adaptive crossover is using uniform scheme, however p_c is not constant during optimization, but it changes base on duration of optimization (6), so for every *i-th* gene crossover probability p_c is calculated. Where $minp_c$ and $.maxp_c$ are min and max vaues of p_c, L is length of chromosome, i_p is number of gene where is $maxp_c$ according to current number of generation G_c and forecasted total number of generations G.

$$p_c(i) = \begin{cases} maxp_c + \dfrac{minp_c - maxp_c}{i_p - 1} \cdot (i-1) & \text{if } i < i_p \\ minp_c + \dfrac{maxp_c - minp_c}{L - i_p} \cdot (i - i_p) & \text{if } i > i_p \\ minp_c & \text{if } i = i_p \end{cases} \qquad (6)$$

$$i_p = \frac{G-1}{G_c - 1} \cdot (L-1) + 1 \qquad (7)$$

This scaling technique is based on hypothesis [19] that in the beginning of the optimization there is necessary to explore solution neighborhood to map it, so probability to exchange gene is higher at the beginning of the chromosome. As optimization continues, $maxp_c$, so i_p is moving towards tail of the chromosome.

4 Experiments and results

The initial testing was done on most the famous Fletcher and Thomson FT10 job shop scheduling problem focusing on minimization of makespan, which is despite its medium combinatorial size (10 jobs on 10 machines with total 100 operations to be scheduled) hard to escape local extremes. Each algorithm has setup with the most close as possible to the original design with exception of mutation which is not applied (see Table 1).

Algortihm	Deafault Population size	Crossover and p_c	parents μ=%λ, [%]	parent selection	Population elimination
SGA	100	uniform 0.85	50	random	μ better individuals survives
BSGA	100	uniform 0.85	50	biased	μ better individuals survives
inGAVAPS	100	one point uniformly random	40	random	Min age=1, Max age=7.
PRofiGA	100, 10/1000 min/max	uniform 0.85	50	tournament	scales acording to the scheme
SAGA	p1=50, p2=100, p3=200	two point uniformly random	50	tournament	scales acording to the scheme
APSC	100, 10/1000 min/max	distributed $minp_c$=0.65, $minp_c$=0.95	50	biased	scales acording to the scheme

Table 1 Algorithm operator setup

Total number of generations is not defined and as Eiben et. al [6] suggests there is defined maximal optimization time. Duration of our experiments is 5 and 15 minutes. This approach should have more comparative value as each algorithm has different population sizing scheme and its size of the population, so total number of generations differs in the same optimization timespan. Figure 1 than present average best makespan approximation of 10 experiments for both 5 and 15 minutes experiments (120 experiments in total) during optimization.

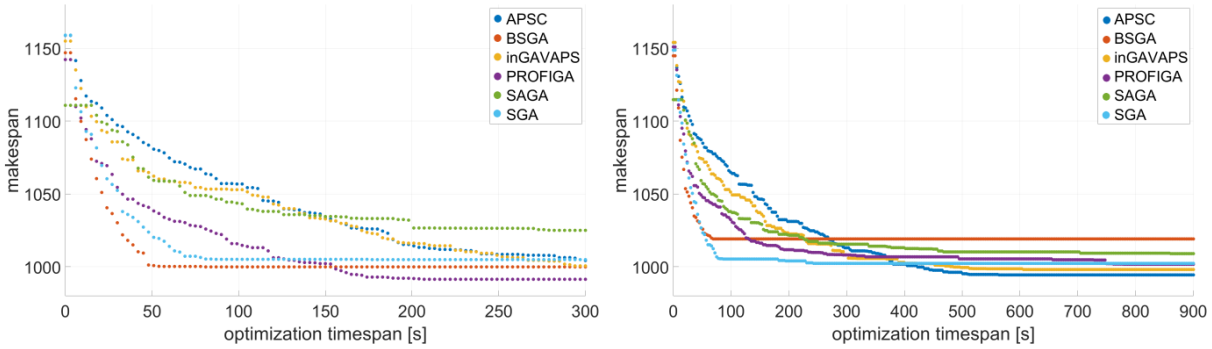


Figure 1 Average makespan progress during optimization of FT10 problem

Results indicate that despite expectation SGA was not outperformed that much by adaptive algorithms. Its progress goes down to optimal solution (optimum - 930) fast, however it stuck in local extreme as expected in less than 100s. BSGA has same approach; however thanks to its biased mechanism progression of makespan was faster with average better results. BSGA has also faster convergence of population to one solution. inGavaPS aging strategy converges much slower to local extreme - in approximately 500s. Its results are in default time not satisfactory in the comparison with previous algorithms. PRofiGA on other hand has continues progression during optimization in both 5 and 15 minutes and has one of the best makespan results. SAGA on the contrary has the worst results. However its progression does not stuck in the local extreme. It is mainly caused by computation demands as there are 3 populations. APSC as expected had very similar optimization progress as PRofiGA and its results are among tested algorithm the most stable comparing best and average best (see Table 2) makespan.

	SGA	BSGA	inGAVAPS	PRofiGA	SAGA	APSC
5 min Average *f(x)*	1004.9	999.9	1013.6	991.4	1033.1	1004.3
5 min Best *f(x)*	979	974	982	948	1006	976
5 min s.deviation *f(x)*	19.10	24.85	17.17	16.92	20.33	17.25
15 min Average *f(x)*	1002.3	1019.1	1050	1001.7	1009.9	994.4
15 min Best *f(x)*	972	988	981	976	958	971
5 min s.deviation *f(x)*	21.44	17.46	24.63	18.41	24.15	14.54

Table 2 Makespan fitness function results of FT10 problem

5 Conclusion

ProfiGA together with APSC are the most promising for real world application as standard deviation are among tested algorithms minimal, so user can expect same results in each run. Additionally, they have also very good best results in terms of objective function as PRofiGA found best solution out of 120 experimental runs (makespan - 948). SAGA algorithm is also very interesting for further use as one of the few algorithms its population does not converges to local extreme and its progress through time suggests that if optimization time was greater, its results would be best.

Generally, presented preliminary results has relatively high level of standard deviation and several cases of 15 minutes experiment runs does not found better solutions in average than 5 minutes experiments. That can be caused by not applying mutation to standard genetic algorithm process (to prevent population convergence) and also by small number of experiment repetition.

These two aspects will be addressed during further research together with more in-depth statistical approach including analysis of genetic operator interaction and influence on the exploration and exploitation properties.

Acknowledgements

This publication was written at Technical University of Liberec as part of the project (21278) – "Optimization of manufacturing systems, 3D technologies and automation" with the support of the Specific University Research Grant, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2018.

References

[1] Arabas, J., Michalewicz, Z., & Mulawka, J. (1994). GAVaPS-a genetic algorithm with varying population size. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence* (pp. 73-78). IEEE.

[2] Bäck, T., Eiben, A. E., & van der Vaart, N. A. (2000, September). An emperical study on GAs "without parameters". In *International Conference on Parallel Problem Solving from Nature* (pp. 315-324). Springer, Berlin, Heidelberg.

[3] Bean, J. C. (1994). Genetic algorithms and random keys for sequencing and optimization. *ORSA journal on computing*, *6*(2), 154-160.

[4] Eggermont, J., & Van Hemert, J. I. (2001). Adaptive genetic programming applied to new and existing simple regression problems. In *European Conference on Genetic Programming* (pp. 23-35). Springer, Berlin, Heidelberg.

[5] Eiben, A. E., Raue, P. E., & Ruttkay, Z. (1994). Genetic algorithms with multi-parent recombination. In *International Conference on Parallel Problem Solving from Nature* (pp. 78-87). Springer, Berlin

[6] Eiben, A. E., & Smit, S. K. (2011). Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm and Evolutionary Computation*, *1*(1), 19-31.

[7] Eiben, A. E., Marchiori, E., & Valko, V. A. (2004, September). Evolutionary algorithms with on-the-fly population size adjustment. In *International Conference on Parallel Problem Solving from Nature* (pp. 41-50). Springer, Berlin, Heidelberg.

[8] Elfeky, E. Z., Sarker, R. A., & Essam, D. L. (2008). Analyzing the simple ranking and selection process for constrained evolutionary optimization. *journal of Computer Science and Technology*, *23*(1), 19-34.

[9] Elsayed, S. M., Sarker, R. A., & Essam, D. L. (2012). The influence of the number of initial feasible solutions on the performance of an evolutionary optimization algorithm. In *Asia-Pacific Conference on Simulated Evolution and Learning* (pp. 1-11). Springer, Berlin, Heidelberg.

[10] Fan, Q., Wang, W., & Yan, X. (2019). Differential evolution algorithm with strategy adaptation and knowledge-based control parameters. *Artificial Intelligence Review*, *51*(2), 219-253.

[11] García-Martínez C., Rodriguez F.J., Lozano M. (2018) Genetic Algorithms. In: Martí R., Pardalos P., Resende M. (eds) *Handbook of Heuristics*. Springer, Cham

[12] Giffler B. & Thompson G. (1960). Algorithms for Solving Production Scheduling Problems. *European Journal of Operational Research*, vol. 8, 487-503.

[13] Gonçalves, J. F., & Resende, M. G. (2018). Random-key genetic algorithms. *Handbook of Heuristics*.

[14] Goldberg, D. E. (2014). Genetic algorithms in search, optimization, and machine learning, addison-wesley, reading, ma, 1989. *NN Schraudolph and J*, *3*(1).

[15] Hinterding, R., Michalewicz, Z., & Peachey, T. C. (1996, September). Self-adaptive genetic algorithm for numeric functions. In *International Conference on Parallel Problem Solving from Nature* (pp. 420-429). Springer, Berlin, Heidelberg.

[16] Jebari, K., Madiafi, M., & Elmoujahid, A. (2013). Parent Selection Operators for Genetic Algorithms. *International Journal of Engineering*, *2*.

[17] Karafotias, G., Hoogendoorn, M., & Eiben, Á. E. (2015). Parameter control in evolutionary algorithms: Trends and challenges. *IEEE Transactions on Evolutionary Computation*, *19*(2), 167-187.

[18] Lin, L., & Gen, M. (2018). Hybrid evolutionary optimisation with learning for production scheduling: state-of-the-art survey on algorithms and applications. *International Journal of Production Research*, *56*(1-2), 193-223.

[19] Manlig, F. (2019). *Optimalizace procesů výukového modelu chytré továrny s principy Průmyslu 4.0*. Liberec. Master thesis. Technická univerzita v Liberci. Supervisor František Koblasa.

[20] Wu, G., Mallipeddi, R., & Suganthan, P. N. (2019). Ensemble strategies for population-based optimization algorithms–A survey. *Swarm and evolutionary computation*, *44*, 695-711.

[21] Raška, P & Zdenek Ulrych.(2018), Testing different evolution strategy selection strategies , *MM Science Journal* , 2290-2299. DOI: 10.17973/MMSJ.2018_03_2017110. ISSN 18031269.

[22] Whitley, L. D. (1989). The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. In Icga (Vol. 89, pp. 116-123).

Semi-parametric spatial model of unemployment dynamics based on Moran's eigenvectors

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Nonlinear relationship among spatially correlated variables, spatially varying relationships, uncontrolled common factors and other sources (potentially simultaneously present) of specification errors can disrupt modelled spatial dependencies if estimated by fully parametric spatial models. A robust semi-parametric model specification and corresponding estimation methods based on Moran's eigenvector maps that address this problem are discussed in our contribution.
Moran's eigenvector maps allow us to extract information from the underlying spatial structure – basically, from a binary connectivity matrix for a relevant sample of geo-coded spatial data. In this type of semi-parametric regression analysis, we work with two distinct types of regressors: a set of geo-coded variables (macroeconomic indicators) and a spatial-filter element. This contribution describes the semi-parametric Moran's eigenvector approach and provides an illustrative estimation of unemployment dynamics in selected EU countries (NUTS2 regional level). Properties of the estimation method and its robustness are discussed.

Keywords: semi-parametric model, spatial analysis, unemployment dynamics.

JEL Classification: C23, C31, C52, E66
AMS Classification: 91B72

1 Introduction

Spatial analysis and "spatially aware" econometric models take into account various types of interaction effects between closely located spatial units (e.g. neighbouring regions). Spatial models have been playing an ever more important role in socio-economic analyses at the regional level. Spatial econometrics uses geo-coded observations of economic variables to estimate quantitative models, evaluate, compare and predict macroeconomic behaviour. As cross-sectional (CS) data are used, there are various model types that can be used for model estimation and evaluation. Besides the widely used (fully) parametric models described in [4], a wide range of non-parametric and semi-parametric estimation approaches are available [5, 9].

In this contribution, we discuss in detail the main principles, features and usage of semiparametric spatial models based on Moran's eigenvectors [6, 11]. An empirical application of the semiparametric model is also provided for illustrative purposes – we evaluate regional unemployment dynamics in 7 EU member states (Belgium, Germany, France, Luxembourg, Netherlands, Portugal and Spain; continental NUTS2 regions, year 2017). Model robustness with respect to changing prior information on spatial structure (neighborhood definition) is also discussed. The remainder of this paper is structured as follows: Section two describes key features and aspects of spatial filters and corresponding semiparametric models, along with references to fundamental literature. Section three provides an illustrative application to the topics outlined. Section four and the list of references conclude our paper.

2 Moran eigenvector maps in spatial regression

Moran's eigenvector maps (MEMs) can be used as a semiparametric spatial filter in different types of regression models (both cross-sectional and panel). Their application is often motivated by circumventing the direct estimation of spatial autoregressive parameters in fully parametric spatial models [7]. Fully parametric models are potentially not robust in a situation where the model suffers from simultaneous presence of different sources of misspecification. Unaccounted nonlinear relationship among spatially correlated variables, spatially varying relationships (non-stationarity), uncontrolled common factors (spatial and time-related) and other instances of spatial heterogeneity can disrupt spatial dependencies. When a MEM is added to regression equation, we work with two distinct types of regressors in our model: a set of conveniently chosen (macroeconomic) indicators following a common spatial structure and a spatial-filter, reflecting main "vectors" (paths) of spatial interaction.

[1] VŠE Praha, nám. W. Churchilla 4 Praha 3, Prague 130 67, Czech Republic, formanek@vse.cz
[2] VŠE Praha, nám. W. Churchilla 4 Praha 3, Prague 130 67, Czech Republic, husek@vse.cz

To produce a MEM for a given sample of observed spatial units, we start by constructing a symmetric connectivity matrix $\boldsymbol{C}$ with binary elements c_{ij} equalling 1 if two spatial units i and j are considered neighbors (e.g. if their spatial distance is below some preset threshold τ). All diagonal elements $c_{ii} = 0$ as spatial units (regions) are not neighbors to themselves by definition. The choice of τ-threshold distance between neighbors has to follow some general rules: all spatial units have to be "connected". Often, τ is chosen using a a minimum spanning tree algorithm – this ensures a path between any two regions. Such path is based on a graph representation of matrix $\boldsymbol{C}$, where spatial units are nodes and edges only exist between neighboring units (if $c_{ij} = 1$). Next, we use $\boldsymbol{C}$ to construct a double-centered connectivity matrix $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = (\boldsymbol{I}_N - \frac{1}{N}\,\boldsymbol{\iota}_N\boldsymbol{\iota}_N')\,\boldsymbol{C}\,(I_N - \frac{1}{N}\,\boldsymbol{\iota}_N\boldsymbol{\iota}_N'), \tag{1}$$

where $\boldsymbol{I}_N$ is the identity matrix (N×N) and $\boldsymbol{\iota}_N$ is a column vector of ones, with length N (the number of spatial units considered). By means of the double-centering, all row and column sums of $\boldsymbol{\Omega}$ equal zero. Moran's eigenvectors $\boldsymbol{v}$ and eigenvalues κ for $\boldsymbol{\Omega}$ are obtained by solving the characteristic equation system $(\boldsymbol{\Omega} - \kappa\boldsymbol{I}_N)\,\boldsymbol{v} = \boldsymbol{0}$. Because $\boldsymbol{\Omega}$ is real and symmetric, all its distinct eigenvectors are orthogonal and linearly independent ($\boldsymbol{\Omega}$ is a rank-deficient matrix). Both positive and negative eigenvalues are produced – eigenvectors corresponding to positive κ_i represent positive spatial association and negative eigenvalues represent negative spatial dependency processes. Spaces spanned by single or multiple $\boldsymbol{v}$ eigenvectors with associated large (positive) eigenvalues κ_i represent global-scale spatial trends (say, landscape-wide core/periphery dynamics in observed EU data). Eigenvectors with medium eigenvalues represent medium scale dynamics (e.g. "regional", say NUTS1 interaction patterns) and eigenvectors with small (positive) eigenvalues would represent small scale dependencies ("local" patchiness, e.g. at the NUTS2 level).

Conveniently chosen $\boldsymbol{v}_i$ eigenvectors – that form a MEM – are included in a regression model. This approach allows us to control for the latent spatial autocorrelation in geo-coded variables. Following a slightly modified methodology described in [11], a relatively simple semiparametric spatial model based on MEMs is derived next. We start by re-casting a spatial Durbin regression model (SDM, see [4] for details and technical discussion) as

$$\boldsymbol{y} = \lambda\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{y}$ is an (N×1) vector of dependent variable observations, $\boldsymbol{W}$ is a row-standardized version of the connectivity matrix $\boldsymbol{C}$ (all row elements in $\boldsymbol{W}$ sum up to 1), $\boldsymbol{X}$ is a matrix of regressors (N×k) – including the intercept column – and $\boldsymbol{\varepsilon}$ is the usual random term. Elements λ, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are parameters of the model (to be estimated). Using a common spatial autocorrelation coefficient $\delta = \lambda$ and assuming $\boldsymbol{\theta} = -\delta\boldsymbol{\beta}$, our SDM specification (2) may be cast as:

$$\boldsymbol{y} = \delta\boldsymbol{W}\boldsymbol{y} + (\boldsymbol{I}_N - \delta\boldsymbol{W})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{3}$$

where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_N)$ and the common factor constraint assumption $\boldsymbol{\theta} = -\delta\boldsymbol{\beta}$ for the term $(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta})$ follows e.g. from [1]. The semiparametric model is established using a misspecification paradigm, where we assume an elementary regression model with spatially autocorrelated disturbances

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*\,, \\ \boldsymbol{\varepsilon}^* &= \boldsymbol{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,, \end{aligned} \tag{4}$$

where $\boldsymbol{\varepsilon}^*$ are the spatially autocorrelated disturbances that may be decomposed into $\boldsymbol{\varepsilon}$ (white noise) and $\boldsymbol{E}$: a set of missing (unobservable, unspecified) exogenous variables that follow a common spatial dependency pattern given by $\boldsymbol{W}$. Finally, $\boldsymbol{\gamma}$ is a vector of parameters.

It is important to note that the misspecification approach to spatial modeling – concentrated in the $\boldsymbol{E}\boldsymbol{\gamma}$ term – is not directly comparable with the preceding specifications (2) and (3). Using the semiparametric approach described next, the misspecification term $\boldsymbol{E}\boldsymbol{\gamma}$ is *approximated* by a set of spatial proxy variables – conveniently chosen MEMs. We start by rewriting the spatial model (3) as

$$\boldsymbol{y} - \delta\boldsymbol{W}\,\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} - \delta\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,. \tag{5}$$

Solving the LHS of (5) for $\boldsymbol{y}$, we get

$$\boldsymbol{y} = (\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}[\boldsymbol{X}\boldsymbol{\beta} - \delta\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}]\,. \tag{6}$$

Using the equivalent expansion $(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} = \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k$, we can cast (6) as

$$\begin{aligned} \boldsymbol{y} &= \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k [\boldsymbol{X}\boldsymbol{\beta} - \delta \boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}], \\ &= \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k (\boldsymbol{X}\boldsymbol{\beta}) - \sum_{k=0}^{\infty} \delta^{k+1} \boldsymbol{W}^{k+1} (\boldsymbol{X}\boldsymbol{\beta}) + \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}, \end{aligned} \tag{7}$$

Because $\delta^0 \boldsymbol{W}^0 = \boldsymbol{I}_N$ and $\sum_{k=0}^{\infty} \delta^{k+1} \boldsymbol{W}^{k+1} = \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k$, we can simplify (7) into

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \left[\sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon} \right] + \boldsymbol{\varepsilon}, \tag{8}$$

which incorporates white noise $\boldsymbol{\varepsilon}$ and the spatial misspecification term, thus following the structure of $\boldsymbol{\varepsilon}^*$ in model (4). Importantly, $\sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}$ is not correlated to the regressors $\boldsymbol{X}$ (by OLS assumption of $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$ independence). Therefore, the OLS-estimated parameters $\hat{\boldsymbol{\beta}}$ of a modified model (4)

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \\ \boldsymbol{\varepsilon}^* &= \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon} \end{aligned} \tag{9}$$

are unbiased estimators for the population parameters $\boldsymbol{\beta}$. On the other hand, the estimated standard errors $s.e.(\hat{\boldsymbol{\beta}})$ will be biased [11]. In empirical applications, the misspecification "problem" as in equation (4) is difficult to deal with because the exogenous term $\boldsymbol{E}\boldsymbol{\gamma}$ is unknown/unspecified/missing. However, taking advantage of the specification (9), we can design spatial proxy variables that satisfy model assumptions: We start by extracting eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ from the quadratic form in

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\} \equiv \text{evec}\left[\boldsymbol{M}_{\boldsymbol{X}}\, \boldsymbol{C}\, \boldsymbol{M}_{\boldsymbol{X}}\right], \tag{10}$$

where $\boldsymbol{M}_{\boldsymbol{X}} = [I_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$ is a projection matrix and the extracted eigenvectors are orthogonal to $\boldsymbol{X}$. Please note that the RHS term in square brackets is a generalization of the centered connectivity matrix introduced in (1). Given the properties of the quadratic form in (10), distinct eigenvectors (of the rank-deficient $\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{C}\boldsymbol{M}_{\boldsymbol{X}}$ matrix) are mutually orthogonal and form a basis for a spatial proxy variable in a semiparametric spatial model. Now, if we follow [11] and associate the term $\boldsymbol{E}$ from (4) with a convenient parsimonious subset of eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$, we can approximate the misspecification term in (8) as

$$\boldsymbol{E}\boldsymbol{\gamma} \approx \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}, \tag{11}$$

where orthogonality to $\boldsymbol{X}$ is maintained. Once the approximate substitution (11) is performed in model (8), the resulting specification may be estimated using the formula

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{12}$$

where $\boldsymbol{y}$ is decomposed into a systematic component (featuring $\boldsymbol{X}$), stochastic spatial component and white-noise residuals. For conveniently specified $\boldsymbol{E}$, the estimated stochastic spatial term $\boldsymbol{E}\hat{\boldsymbol{\gamma}}$ removes a significant portion of the mean squared error (MSE) term attributable to spatial autocorrelation (i.e. $\boldsymbol{E}\hat{\boldsymbol{\gamma}}$ is often referred to as spatial filter). Overall, Tiefelsdorf and Griffith [11] conclude that this filtering approach is fairly robust to model specification errors when compared with fully parametric models and ML-based estimators.

Choosing a convenient and parsimonious subset of eigenvectors for the $\boldsymbol{E}$ term is crucial for a successful application of the MEM-based semiparametric algorithm. We aim to choose such $\boldsymbol{E}$ so that the residuals $\hat{\boldsymbol{\varepsilon}}$ from model (12) become spatially random (independent with respect to the underlying spatial domain). Also, we aim to find a parsimonious, (i.e. smallest possible) subset of eigenvectors leading to spatial independence of $\hat{\boldsymbol{\varepsilon}}$. To establish the subset of eigenvectors $\boldsymbol{E}$ with desired properties, stepwise regression approach is often suggested, based on a modified Moran's I coefficient [1], often denoted [11] as MC and formalized as

$$MC_{\boldsymbol{v}_i} = \frac{N}{\boldsymbol{\iota}_N' \boldsymbol{Z}\, \boldsymbol{\iota}_N} \boldsymbol{v}_i' \boldsymbol{W}\, \boldsymbol{v}_i, \tag{13}$$

where $\boldsymbol{Z}$ is a distance-based similarity matrix with $z_{ij} = 1 - (\frac{h_{ij}}{\max(h_{ij})})^2$. For conveniently measured distances h_{ij} between units i and j (e.g. in km), individual z_{ij} values vary between zero for $h_{ij} = \max(h_{ij})$ and 1 for $h_{ij} = 0$. The expression (13) is a coefficient that is commonly provided and cited in literature covering semiparametric MEM-based models. The first part of the RHS is just a scaling element (the fraction is fixed for a given sample of data). Hence, in the stepwise algorithm described next, we may simply focus on the $\boldsymbol{v}_i' \boldsymbol{W} \boldsymbol{v}_i$ element of (13).

A forward stepwise selection method by Griffith's [6] may be described as follows: The first eigenvector $\boldsymbol{v}_1$ is chosen based on maximizing $MC_{\boldsymbol{v}_i}$ in expression (13). Using such $\boldsymbol{v}_1$ as a starting eigenvector subset for $\boldsymbol{E}$, equation (12) is estimated and corresponding residuals $\hat{\boldsymbol{\varepsilon}}$ are evaluated with respect to their spatial autocorrelation (e.g. using Moran's I statistic [8]). If residuals are spatially dependent, new eigenvector, $\boldsymbol{v}_2$ is added to $\boldsymbol{E}$ using the same MC-maximization criterion and spatial autocorrelation in residuals is tested again. Eigenvectors are iteratively added to $\boldsymbol{E}$, until spatial autocorrelation in residuals $\hat{\boldsymbol{\varepsilon}}$ falls below a predetermined threshold (say, until we fail to reject the null hypothesis of no spatial autocorrelation at the 5 % significance level). Eigenvectors in $\boldsymbol{E}$ are mutually orthogonal and follow a strictly decreasing sequence, where each eigenvector explains a specific proportion of variance in residuals of the model (12) – the largest proportion of variance is explained by the first eigenvector selected into $\boldsymbol{E}$, the second largest amount of variance is explained by the second eigenvector, etc. Computational efficiency and improvements to the stepwise algorithm are discussed e.g. by Tiefelsdorf and Griffith [11].

The above discussed search methods for eigenvector components of $\boldsymbol{E}$ focus on minimizing spatial autocorrelation in $\hat{\boldsymbol{\varepsilon}}$. However, this is not the only possible paradigm and other approaches are available from literature as well [7]. For example, we can disregard spatial autocorrelation in residuals of model (12) and use stepwise (or exhaustive / brute force) approach for selection of individual eigenvectors in $\boldsymbol{E}$ to minimize the total variance in residuals of the model (12).

3 Illustrative empirical application: regional unemployment dynamics

As outlined in Section 1, an empirical application of a semiparametric MEM-based model is provided next. Using an economically diverse yet relatively spatially compact sample of 103 NUTS2 regions of Belgium, Germany, France, Luxembourg, Netherlands, Portugal and Spain (all islands and overseas regions are excluded to avoid "breaks" in spatial structure – those are incompatible with the analysis framework), model (14) is estimated as

$$Unem_i = \beta_0 + \beta_1 \log(GDPpc_i) + \beta_2 CapFor_i + \boldsymbol{s}_i\boldsymbol{\delta} + \boldsymbol{e}_i\boldsymbol{\gamma} + \varepsilon_i\,, \tag{14}$$

where $Unem_i$ is the unemployment rate in a given ith NUTS2 region, $\log(GDPpc_i)$ is the log-transformed GDP per capita (in EUR), $CapFor_i$ is a relative indicator of capital formation, defined as a proportion of total capital formation to GDP. While 2017 data are used for the first two variables described, capital formation is calculated using 2016 data ($t-1$). This approach accounts for the empirically established time lag between capital formation expenditures and actual effects to the labor market. The element $\boldsymbol{s}_i$ denotes a vector of dummy variables identifying the countries included in our sample (using this approach, individual fixed-effect intercepts are estimated for each country); $\boldsymbol{s}_i$ is a 6×1 row-vector and Belgium is used as a reference. Vector $\boldsymbol{e}_i$ is the ith row of matrix $\boldsymbol{E}$ as described in equation (12). Elements $\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2)'$, $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are parameters of the model to be estimated and ε_i is a random term (properties discussed in Section 2).

Regressors	**Coefficients**	**Std. errs.**	***t*-values**	***p*-values**
(Intercept)	48.404	9.337	5.186	0.000
log(*GDPpc*)	-3.394	0.873	-3.887	0.000
CapFor	-23.363	7.055	-3.312	0.001
s_{DE}	-4.100	0.666	-6.158	0.000
s_{ES}	6.487	0.837	7.750	0.000
s_{FR}	0.698	0.719	0.970	0.335
s_{LU}	0.255	2.119	0.120	0.904
s_{NL}	-2.616	0.787	-3.323	0.001
s_{PT}	-2.883	1.287	-2.239	0.028

Table 1 Estimation output – unemployment dynamics, 2017 data

All data are obtained from Eurostat. For reproducibility purposes, names of individual (source) datasets are given: "lfst_r_lfu3rt" covers unemployment data, "nama_10r_2gdp" provides GDP information (per capita and

total per region) and "nama_10r_2gfcf" is used to retrieve capital formation data. State level information was generated from the NUTS2 region ID (where the first two digit identify a country, i.e. NUTS0 region). Using model (14) and the data described here, model parameters were estimated using specialized R packages [2, 10], designed for handling spatial data estimation of MEM-based semiparametric models. Estimation results are shown in Table 1.

From the estimation output as provided in Table 1, we may see that the estimated β_j coefficients follow both macroeconomic theory as well as general empirical findings from other studies [3, 5]. There is a clear and statistically significant inverse relation between GDP growth and unemployment rates and the same type of ceteris paribus effect may be observed for the capital formation variable as used in model (14). Also, the δ-coefficients corresponding to country-specific dummy variables stacked in vector $\boldsymbol{s}$ indicate the existence of pronounced differences across labor markets in individual economies (states) included in our sample. This conclusion can be made despite of the statistically insignificant coefficients for France and Luxembourg (all other economies differ significantly from the reference level at $\alpha = 0.05$).

The estimated $\boldsymbol{\gamma}$ coefficients are left out from Table 1 due to space constraints for the contribution and given their limited interpretation values. However, Figure 1 is presented for reader's convenience as it may provide useful insight into the estimation of MEM-based spatial models. Figure 1 contains all nine eigenvectors (columns of matrix $\boldsymbol{E}$) used in estimation of model (14) – as selected by the stepwise algorithm based on MC coefficient (13). From Figure 1, we can see how the v_1, v_2, v_3 eigenvectors reflect area-wide patterns (say, East-West) and subsequent $\boldsymbol{E}$ elements reflect smaller-scale patterns. The eigenvectors shown in Figure 1 correspond to a spatial structure (neighborhood definition) based on a maximum distance threshold $\tau = 250$ km that applies to the individual c_{ij} elements of $\boldsymbol{C}$ as used in expression (10).

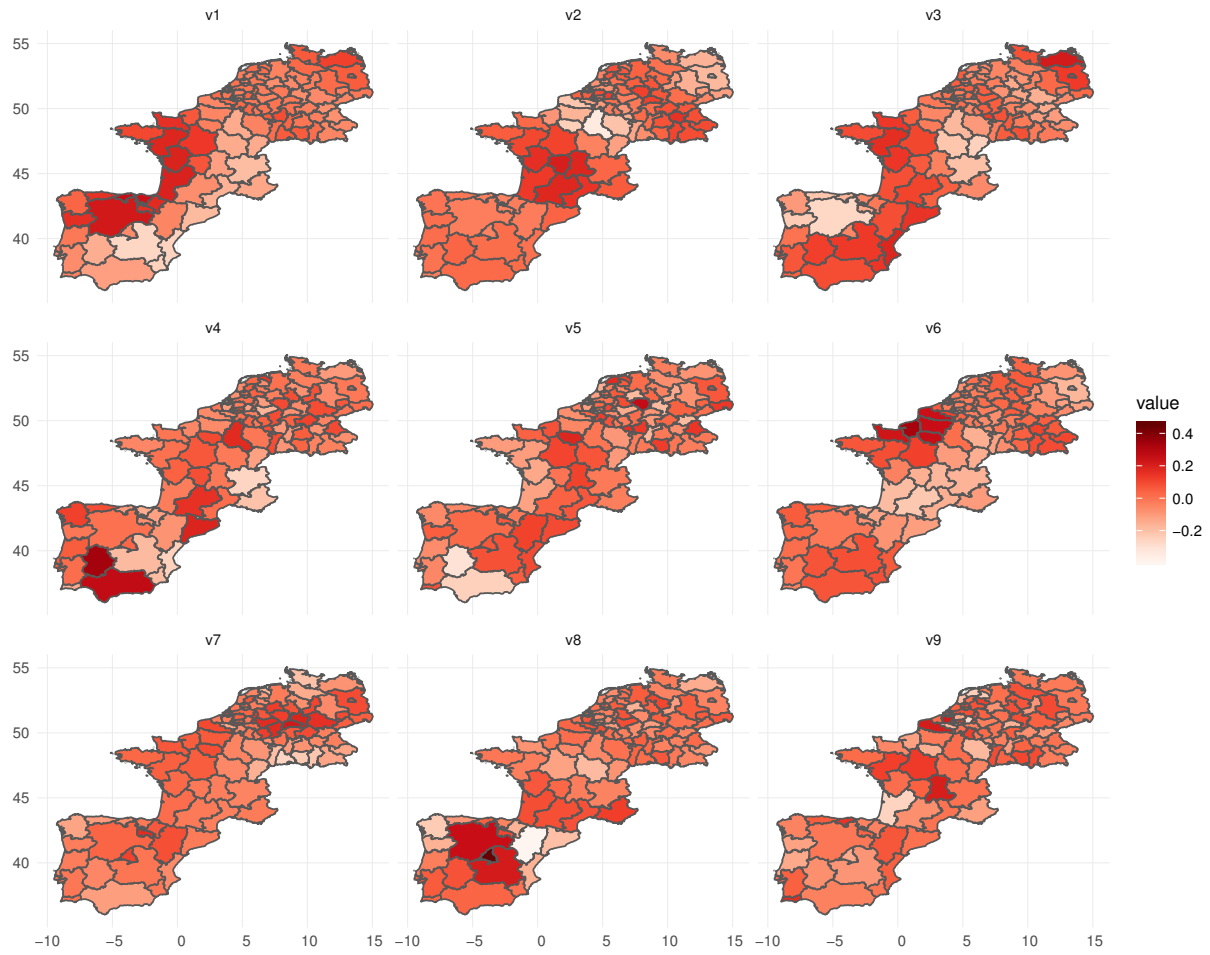


Figure 1 Moran's eigenvector map (MEM) used in regression model

The residuals of model (14) are spatially random – hence, our model adequately explains all spatial interactions. Please note that if τ is changed, the eigenvectors generated through expression (10) and the corresponding stepwise

selection would yield different results. While the estimated parameters of model (12) or its empirical application (14) are not directly affected (as $\boldsymbol{E}$ elements are orthogonal to $\boldsymbol{X}$ by construction), overall model performance (predictive properties, variance of residuals, etc.) may be affected. Hence, in spatial analysis, one should always evaluate model robustness with respect to changes in spatial structure (this applies to both fully parametric and semi-parametric spatial models [5]).

4 Conclusions

Spatial filtering and semi-parametric econometric models provide a versatile framework for quantitative analyses in many fields of economic research. This contribution focuses on methodology of a specific type of semi-parametric spatial models: Moran's eigenvectors are described and their use in econometric models is discussed and evaluated. Also, an empirical illustrative analysis of unemployment is provided. Our results underline the importance of controlling for spatial dependency in macroeconomic analyses at the regional level. Also, the above described approach might be extended to a panel data analysis framework and/or used in similar types of analyses.

While this paper focuses on unemployment dynamics in selected EU economies, diverse aspects of regional interactions are present in most macroeconomic variables and processes – thus, their study and explicit incorporation in econometric models is a widely applicable approach.

Acknowledgements

This work is supported through grant IGA F4/60/2018 from the Faculty of Informatics and Statistics, University of Economics, Prague. Geo-data source: GISCO-Eurostat (European Commission), Administrative boundaries: © EuroGeographics.

References

[1] Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht: Kluwer.

[2] Bivand, R. & Piras, G. (2015). Comparing iplementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63(18), 1–36.

[3] Cuestas, J. C., & Gil-Alana, L. A. (2017). Unemployment rate cycles in Europe. *Applied Economics Letters*, 24(2), 136–139.

[4] Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. New York: Springer.

[5] Formánek, T. & Hušek, R. (2018). Getis' non-parametric approach to spatial data analysis, with application to regional unemployment dynamics. In L. Váchová, V. Kratochvíl (Eds.), *Mathematical Methods in Economics (MME2018)* (pp. 92–97). MatfyzPress, Publishing House of the Faculty of Mathematics and Physics Charles University.

[6] Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Berlin: Springer.

[7] Griffith, D. A., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, 87(10), 2603–2613.

[8] LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press, Taylor & Francis Group.

[9] Patuelli, R. & Griffith, D. A. & Tiefelsdorf, M. Nijkamp, P. (2011). Spatial filtering and eigenvector stability: space-time models for German unemployment data. *International Regional Science Review*, 34(2), 253–280.

[10] Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, 10(1), 439–446.

[11] Tiefelsdorf, M. & Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* 39(5), 1193–1221.

Jensen Alpha for oriented fuzzy discount factor

Anna Łyczkowska-Hańćkowiak[1]

Abstract. The analysis presented in this paper regards the security of a present value given as an ordered fuzzy number. The present value was estimated in an imprecise manner and supplemented by the forecast of its coming changes. A discount factor of such security is an ordered fuzzy number of the orientation identical to the oriented present value that determines it. All classical methods of portfolio analysis are based on the definition of the return rate. In the case of securities with a fuzzy present value, a discount factor is a better tool for portfolio analysis than the return rate, which implies the chosen methods of management of securities should be revised and transformed to equivalent methods based on a discount factor. This would enable the use of those methods in the case of a financial instrument of the oriented fuzzy present value. This paper presents example results of the realization of such a postulate. The main aim of the paper is to generalize Jensen's ratio to a case of investment recommendations management formulated for a security characterized by an oriented discount factor. A five-degree rating scale was used. The whole deliberation is illustrated by broad numerical examples.

Keywords: Jensen's ratio, ordered fuzzy number, fuzzy oriented discount factor.

JEL Classification: C52, C61, C65
AMS Classification: 03E72, 06A06, 90C70

1 Introduction

Present value (PV) is defined in [14] as a present equivalent of a payment available in a given time in the future. PV of future cash flows is an approximate value, with fuzzy numbers being one of the main tools of its modelling. If present value is evaluated by a fuzzy number, then the return rate is considered as a fuzzy probabilistic set [13]. The expected return rate is obtained as a fuzzy subset in the real line. This result is a theoretical foundation for investment strategies presented in [15, 18]. In [16] these results are generalized for the case when the present value is evaluated by intuitionistic fuzzy number [1]. Ordered fuzzy numbers (OFN) are defined in an intuitive way by Kosiński and his collaborators [4-7]. The competent elaboration on the current state-of-art on OFNs is the monograph [21]. Kosiński has shown that there exist such OFNs which are not fuzzy numbers [4]. For this reason, the original Kosiński's theory was revised in [17].

The main aim of the paper is the analysis of the possibility to expand the investment strategies to a case when the present value is examined via ordered fuzzy numbers. To fulfil that task, Jensen's ratio [3] will be extended for that case. In the original Jensen's ratio, the main premise for formulating investment recommendation is the expected return rate of the analysed security. In [19, 20] it is shown that the expected fuzzy discount factor is a better tool for appraising considered securities than the fuzzy expected return rate. It was shown that the use of the expected fuzzy discount factor significantly facilitates the portfolio analysis. For this reason, the original Jensen's ratio will be transformed to an equivalent form in which a basic premise to form investment recommendation is the expected fuzzy discount factor in case of analysed securities.

2 Ordered fuzzy numbers – basic facts

An imprecise number is a family of values in which each considered value belongs to it in a varying degree. A commonly accepted model of the imprecise number is the fuzzy number (FN), defined as a fuzzy subset of the real line $\mathbb{R}$. The most general definition of FN and arithmetic operations on FN which are coherent with the Zadeh Extension Principle were given by Dubois and Prade [2]. Ordered fuzzy numbers (OFN) were intuitively introduced by Kosiński and his co-writers in the series of papers [4-7] as an extension of the concept of FN. A significant drawback of Kosiński's theory is that there exist OFNs which are not FN [4]. The OFNs' usefulness results from the fact that an OFN is defined as FN supplemented by its orientation determined below Defi-

[1] WSB University in Poznań, Institute of Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, anna.lyczkowska-hanckowiak@wsb.poznan.pl

nition 1. The FN orientation is interpreted as prediction of future FN changes. The Kosiński's theory was revised by Piasecki in [17]. We will limit our deliberations to a special case of ordered fuzzy numbers – trapezoidal ordered fuzzy numbers defined in [17] in a following manner.

Definition 1. For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$ the trapezoidal ordered fuzzy number (TrOFN) $\overrightarrow{Tr}(a,b,c,d)$ is defined as the pair of FNs determined by their membership function $\mu_{\overrightarrow{Tr}}(\cdot\,|a,b,c,d) \in [0;1]^{\mathbb{R}}$ given by orientation $[\![a \rightarrowtail d]\!]$ and the identity

$$\mu_{\overrightarrow{Tr}}(x|a,b,c,d) = \begin{cases} 0, & x \notin [a,d] = [d,a], \\ \frac{x-a}{b-a}, & x \in [a,b[=]b,a], \\ 1, & x \in [b,c] = [c,b], \\ \frac{x-d}{c-d}, & x \in]c,d] = [d,c[. \end{cases} \tag{1}$$

The space of all TrOFNs will be denotes by the symbol $\mathbb{K}_{Tr}$. The condition $a < d$ fulfilment determines the positive orientation $[\![a \rightarrowtail d]\!]$ of TrOFN $\overrightarrow{Tr}(a,b,c,d)$, which is interpreted as such an imprecise number, which may increase. The condition $a > d$ fulfilment determines the negative orientation of TrOFN $\overrightarrow{Tr}(a,b,c,d)$, which is interpreted as such an imprecise number, which may decrease. For the case $a = d$, TrOFN $\overrightarrow{Tr}(a,b,c,d)$ represents a crisp number $a \in \mathbb{R}$, which is not oriented.

Kosiński has introduced the arithmetic operators of dot product $\odot$ for any real number $\beta \in \mathbb{R}$ and any TrOFN $\overrightarrow{Tr}(a,b,c,d)$ as follows:

$$\beta \odot \overrightarrow{Tr}(a,b,c,d) = \overrightarrow{Tr}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d). \tag{2}$$

The space of ordered fuzzy numbers is not closed under Kosiński's addition. Therefore, the Kosinski's theory is modified in this way that the space of ordered fuzzy numbers is closed under revised arithmetic operations. In case of any TrOFNs $\overrightarrow{Tr}(a,b,c,d)$ and $\overrightarrow{Tr}(p-a,q-b,r-c,s-d)$ their sum is determined as follows [17]:

$$\begin{aligned} &\overrightarrow{Tr}(a,b,c,d) \boxplus \overrightarrow{Tr}(p-a,q-b,r-c,s-d) = \\ &= \begin{cases} \overrightarrow{Tr}(\min\{p,q\},q,r,\max\{r,s\}), & (q<r) \vee (q=r \wedge p \le s), \\ \overrightarrow{Tr}(\max\{p,q\},q,r,\min\{r,s\}), & (q>r) \vee (q=r \wedge p > s). \end{cases} \end{aligned} \tag{3}$$

In the set of trapezoidal ordered fuzzy numbers, the relation of a fuzzy preorder [18] was determined. Let us consider the pair $(\overrightarrow{\mathcal{K}},\overrightarrow{\mathcal{L}}) \in \mathbb{K}_{Tr} \times \mathbb{K}_{Tr}$. On the set $\mathbb{K}_{Tr}$ of all OFNs we define the relation $\overrightarrow{\mathcal{K}} \succcurlyeq \overrightarrow{\mathcal{L}}$ as follows:

$$\overrightarrow{\mathcal{K}} \succcurlyeq \overrightarrow{\mathcal{L}} \Leftrightarrow \text{“}TrOFN\ \overrightarrow{\mathcal{K}}\ is\ greater\ or\ equal\ to\ TrOFN\ \overrightarrow{\mathcal{L}}.\text{”} \tag{4}$$

This relation is a fuzzy preorder $Q \in \mathcal{F}(\mathbb{K}_{Tr} \times \mathbb{K}_{Tr})$ determined by means of such membership function $\nu_Q \in [0,1]^{\mathbb{K}_{Tr} \times \mathbb{K}_{Tr}}$ that from the point of view of multivalued logic, the value $\nu_Q(\overrightarrow{\mathcal{K}},\overrightarrow{\mathcal{L}})$ may be interpreted as true-value of the sentence (4). The variability of membership function ν_Q is described in detail as follows:

Theorem 1. For any pair $(\overrightarrow{\mathcal{K}},\overrightarrow{\mathcal{L}}) \in \mathbb{K}_{Tr} \times \mathbb{K}_{Tr}$ satisfying the condition

$$\overrightarrow{\mathcal{K}} \boxplus \left((-1) \odot \overrightarrow{\mathcal{L}}\right) = \overrightarrow{\mathcal{M}} = \overrightarrow{Tr}(a_{\mathcal{M}}, b_{\mathcal{M}}, c_{\mathcal{M}}, d_{\mathcal{M}}) \tag{5}$$

we have:

– if $a_{\mathcal{M}} \le d_{\mathcal{M}}$, then $\nu_Q(\overrightarrow{\mathcal{K}},\overrightarrow{\mathcal{L}}) = \begin{cases} 0, & 0 > d_{\mathcal{M}}, \\ \frac{-d_{\mathcal{M}}}{c_{\mathcal{M}} - d_{\mathcal{M}}}, & d_{\mathcal{M}} \ge 0 > c_{\mathcal{M}}, \\ 1, & c_{\mathcal{M}} \ge 0, \end{cases}$ (6)

– if $a_{\mathcal{M}} > d_{\mathcal{M}}$, then $\nu_Q(\overrightarrow{\mathcal{K}},\overrightarrow{\mathcal{L}}) = \begin{cases} 0, & 0 > a_{\mathcal{M}}, \\ \frac{-a_{\mathcal{M}}}{b_{\mathcal{M}} - a_{\mathcal{M}}}, & a_{\mathcal{M}} \ge 0 > b_{\mathcal{M}}. \\ 1, & b_{\mathcal{M}} \ge 0. \end{cases}$ □ (7)

3 Oriented fuzzy present value

The present value (PV) is called the current equivalent value of payments at a fixed point in time [14]. PV of the future cash flow may be imprecise. For this reason, the PV is described using fuzzy numbers. Then PV is characterized by a monotonic sequence $\{V_s, V_f, \check{C}, V_l, V_e\}$, where $\check{C}$ is a market price, $[V_s, V_e] \subset \mathbb{R}^+$ is an interval of all possible PV values, $[V_f, V_l] \subset [V_s, V_e]$ is an interval of all prices which do not perceptibly differ from the market price $\check{C}$. PV was estimated in an imprecise manner and it was supplemented by a forecast of its closest changes. Such a present value is called the oriented present value (OPV). OPV is estimated by TrOFN:

$$\overrightarrow{PV} = \overrightarrow{Tr}(V_s, V_f, V_l, V_e). \tag{8}$$

If we predict a rise in the market price, then OPV is described by a positively oriented TrOFN. If we predict a fall in market price, then OPV is described by a negatively oriented TrOFN. If $V_s < V_e$, then the positively oriented PV means the forecast of the value increase. If $V_s < V_e$, then the negative orientation is the forecast of the decrease in value. In [10], shares' OPVs are determined as TrOFNs describing their Japanese candles [12].

Example 1 [10]: We evaluate the portfolio π composed of blocks of shares in Assecopol (ACP), ENERGA (ENG), JSW (JSW), KGHM (KGH), LOTOS (LTS), ORANGEPL (OPL) and PKOBP (PKO). Based on closing of the session on the Warsaw Stock Exchange on January 15, 2018, for each considered share we determine its OPV as TrOFN describing its Japanese candle. Obtained in this way shares' OPVs are presented in Table 1. For each portfolio component, we determine its market price $\check{C}_s$ as initial price on 16.01.2018.

Stock company	Present value $\overleftrightarrow{PV}_s$	Market price $\check{C}_s$	Expected return rate $\bar{r}_s$	CAPM factor β_s
ACP	$\overleftrightarrow{Tr}(45.90;\ 45.90;\ 45.50;\ 45.48)$	45.70	0.0300	1.95
CPS	$\overleftrightarrow{Tr}(22.92;\ 22.82;\ 22.82;\ 22.76)$	22.82	0.0355	2.04
ENG	$\overleftrightarrow{Tr}(10.22;\ 10.19;\ 10.17;\ 10.14)$	10.18	0.0150	0.98
JSW	$\overleftrightarrow{Tr}(92.24;\ 92.54;\ 92.54;\ 92.80)$	92.54	0.0400	2.03
KGH	$\overleftrightarrow{Tr}(102.65;\ 103.05;\ 103.60;\ 103.90)$	103.33	0.0390	2.94
LTS	$\overleftrightarrow{Tr}(56.70;\ 56.56;\ 56.40;\ 56.28)$	56.48	0.0450	2.00
OPL	$\overleftrightarrow{Tr}(5.75;\ 5.76;\ 5.90;\ 5.90)$	5.83	0.0360	2.28
PGE	$\overleftrightarrow{Tr}(10.39;\ 10.39;\ 10.35;\ 10.33)$	10.37	0.0235	1.28
PKO	$\overleftrightarrow{Tr}(42.61;\ 42.61;\ 43.22;\ 43.22)$	42.91	0.0420	2.76

Table 1 Evaluation of portfolio π components. Source [10] & own elaboration

We notice that the stock companies KGH, JWS, OPL and PKO are evaluated by positively oriented PV, which predicts a rise in market price. Similarly the stock companies ACP, CPS, ENG, LTS and PGE are evaluated by negatively oriented PV, which predicts a fall in market price.

4 Oriented fuzzy discount factor

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the security considered here is determined by two values: anticipated $FV = V_t$ and assessed $PV = V_0$. The basic characteristic of benefits from owning this security is the simple return rate defined as: $r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1$.

In practice of financial markets analysis, the uncertainty risk is usually described by probability distribution of return rate calculated for $V_0 = \check{C}$. After Markowitz [11] we assume that this simple return rate has the Gaussian distribution $N(\bar{r}, \sigma)$. Then the expected discount factor (EDF) $\bar{v} \in \mathbb{R}$ is given by the identity $\bar{v} = \frac{1}{1+\bar{r}}$.

Example 2: All considerations in the paper are run for the quarterly period of investment time $t = 1\ quarter$. Expected return rates of portfolio components π are presented in Table 1.

Oriented expected fuzzy discount factor (OEDF) described by TrOFN was given in [9]

$$\overleftrightarrow{V} = \overleftrightarrow{Tr}\left(\frac{V_s}{\check{C}} \cdot \bar{v}, \frac{V_f}{\check{C}} \cdot \bar{v},\ \frac{V_l}{\check{C}} \cdot \bar{v}, \frac{V_e}{\check{C}} \cdot \bar{v}\right). \tag{9}$$

Example 3: Using (9), we calculate quarterly EDF and OEDF for each component of the portfolio π. Obtained in this way evaluations are presented in Table 2.

The discount factor of a security described in this way is an oriented fuzzy number with the identical orientation as the oriented present value used for estimation. It is worth stressing that the maximum criterion of the expected return rate can be replaced by the minimum criterion of the expected discount factor. In case of fuzzy values of both parameters, those criteria are equivalent.

5 Investment recommendations

The investment recommendation is the counsel given by the advisors to the investor. In this paper we will consider the family of advice which is applied in [15]. Therefore, recommendations are expressed by means of standardized advice [15]. The set $\mathbb{A} = \{A^{++},\ A^{+},\ A^{0},\ A^{-},\ A^{--}\}$ - called a rating scale, constitute advices, where

– A^{++} denotes the advice Buy, i.e. suggesting that evaluated security is significantly undervalued,
– A^{+} denotes the advice Accumulate, i.e. suggesting that evaluated security is undervalued,
– A^{0} denotes the advice Hold, i.e. suggesting that evaluated security is fairly valued,
– A^{-} denotes the advice Reduce, i.e. suggesting that evaluated security is overvalued,
– A^{--} denotes the advice Sell, i.e. suggesting that evaluated security is significantly overvalued.

Due to such approach we will be able to compare the obtained results with the results of similar research conducted in [15, 16, 19].

Let us take into account a fixed security $\check{S}$, represented by the pair $(\bar{r}_s,\ \varpi_s)$ where $\bar{r}_s$ is an expected return on $\check{S}$ and ϖ_s is a parameter characterizing the security $\check{S}$. Adviser's counsel depends on the expected return. The criterion for a competent choice of advice can be presented as a comparison of the values profit $g(\bar{r}_s|\varpi_s)$ and the profitability threshold (PT) $\check{G}$, where $g(\cdot\,|\varpi_s)\colon \mathbb{R}\to\mathbb{R}$ is an increasing function of the expected return rate. By the symbol $\mathbb{S}$ we denote the set of all considered securities. The advice choice function $\Lambda\colon \mathbb{S}\times\mathbb{R}\to 2^{\mathbb{A}}$ was given in [15]. This way was assigned the advice subset $\Lambda(\check{S},\check{G})\subset\mathbb{A}$ which is interpreted as the investment recommendation given for the security. We can assume that a given security $\check{S}$ is represented by the ordered pair $(\bar{v}_s,\ \varpi_s)$ where $\bar{v}_s=\frac{1}{1+\bar{r}_s}$ is EDF for $\check{S}$. It is very easy to check that we have

$$g(\bar{r}_s|\varpi_s)\geq\check{G}\Leftrightarrow\bar{v}_s\leq\frac{1}{1+g^{-1}(\check{G}|\varpi_s)}=H_s, \tag{10}$$

$$g(\bar{r}_s|\varpi_s)\leq\check{G}\Leftrightarrow\bar{v}_s\geq H_s. \tag{11}$$

The value H_s is interpreted as a specific profitability threshold (SPT) determined for the security $\check{S}$. Then advice choice function $\Lambda\colon \mathbb{S}\times\mathbb{R}\to 2^{\mathbb{A}}$ is equivalently described in a following way

- $A^{++}\in\Lambda(\check{S},\check{G})\Leftrightarrow\bar{v}_s\leq H_s\wedge\neg\bar{v}_s\geq H_s$,
- $A^{+}\in\Lambda(\check{S},\check{G})\Leftrightarrow\bar{v}_s\leq H_s$,
- $A^{0}\in\Lambda(\check{S},\check{G})\Leftrightarrow\bar{v}_s\leq H_s\wedge\ \bar{v}_s\geq H_s$, (12)
- $A^{-}\in\Lambda(\check{S},\check{G})\Leftrightarrow\bar{v}_s\geq H_s$,
- $A^{--}\in\Lambda(\check{S},\check{G})\Leftrightarrow\neg\bar{v}_s\leq H_s\wedge\ \bar{v}_s\geq H_s$.

We consider the case when EDFs are imprecisely evaluated. Moreover, we can additively predict future changes in EDF value. Then a given security $\check{S}$ is represented by the ordered pair $(\vec{\mathcal{V}}_s,\ \varpi_s)$ where

$$\vec{\mathcal{V}}_s=\overrightarrow{Tr}\left(D_b^{(s)},D_f^{(s)},D_l^{(s)},D_b^{(s)}\right) \tag{13}$$

is OEDF calculated with use (9) for $\check{S}$. For this case, using (10) we calculate specific profitability threshold H_s. If the PT $\check{G}$ is given, then each SPT H_s may be represented by TrOFN

$$\overrightarrow{[\![H_s]\!]}=\overrightarrow{Tr}(H_s,H_s,H_s,H_s). \tag{14}$$

Then the value $\tilde{\Lambda}(\check{S},\check{G})$ of the recommendation choice function $\tilde{\Lambda}\colon[0,1]^{\mathbb{R}}\times\mathbb{R}\to[0,1]^{\mathbb{R}}$ is the membership function $\lambda(\cdot\,|\check{S},\check{G})\colon\mathbb{A}\to[0,1]$ function determined in accordance with (12) in the following way:

- $\lambda(A^{++}|\check{S},\check{G})=\nu_Q\left(\overrightarrow{[\![H_s]\!]},\vec{\mathcal{V}}_s\right)\ \wedge\ \left(1-\nu_Q\left(\vec{\mathcal{V}}_s,\overrightarrow{[\![H_s]\!]}\right)\right)$,
- $\lambda(A^{+}|\check{S},\check{G})=\nu_Q\left(\overrightarrow{[\![H_s]\!]},\vec{\mathcal{V}}_s\right)$,
- $\lambda(A^{0}|\check{S},\check{G})=\nu_Q\left(\overrightarrow{[\![H_s]\!]},\vec{\mathcal{V}}_s\right)\ \wedge\ \nu_Q\left(\vec{\mathcal{V}}_s,\overrightarrow{[\![H_s]\!]}\right)$, (15)
- $\lambda(A^{-}|\check{S},\check{G})=\nu_Q\left(\vec{\mathcal{V}}_s,\overrightarrow{[\![H_s]\!]}\right)$,
- $\lambda(A^{--}|\check{S},\check{G})=\nu_Q\left(\vec{\mathcal{V}}_s,\overrightarrow{[\![H_s]\!]}\right)\ \wedge\ \left(1-\nu_Q\left(\overrightarrow{[\![H_s]\!]},\vec{\mathcal{V}}_s\right)\right)$.

From the point of view of a multivalued logic, the value $\lambda(A|\check{S},\check{G})$ is interpreted as a logic value of the sentence

$$"Recommendation\ A\in\mathbb{A}\ is\ advised\ to\ take\ an\ investment\ decision". \tag{16}$$

From the point-view of decision-making, the value $\lambda(A|\check{S},\check{G})$ is interpreted as a degree of recommendation support $A\in\mathbb{A}$, i.e. a declared share of the advisor in the responsibility in case of final decision-making according to the advice $A\in\mathbb{A}$. In the described situation the investment recommendation $\tilde{\Lambda}(\check{S},\check{G})$ is the fuzzy subset in the rating scale $\mathbb{A}$.

6 Jensen's Alpha

Jensen's Alpha is one of the criteria of risk management. In this model of financial equilibrium, the compared values are the expected return on a security and the directional factor of the CAPM model assigned to this security. Jensen's Alpha estimates amount of the premium for market risk . On the capital market we observe the

risk-free return r_0 and the expected market return r_M. The security $\check{S}$ is represented by the pair (r_s, β_s), where β_s is the directional factor of the CAPM model assigned to this instrument.

Example 4. We focus on the Warsaw Stock Exchange. We consider financial market with risk-free bound instrument determined as quarterly treasure bounds with return rate $r_0 = 0.0075$. The market portfolio is represented by portfolio determining stock exchange index WIG20 which have expected return rate $r_M = 0.0200$. The CAPM directional factors for each portfolio component are presented in Table 1.

If the security $\check{S}$ is represented by the pair $(\bar{r}_s, \beta_s)$, then defined by Jensen, the profit index $g(\cdot|\sigma_s)\colon \mathbb{R} \to \mathbb{R}$ and the PT $\check{G}$ are defined as follows:

$$g(\bar{r}_s|\beta_s) = r_s - \beta_s(r_M - r_0), \tag{17}$$

$$\check{G} = r_0. \tag{18}$$

For this case, we calculate SPT H_s in following manner

$$H_s = \frac{1}{1+r_0+\beta_s(r_M-r_0)}. \tag{19}$$

Example 5: Using (19), we calculate SPT for each component of the portfolio π. Evaluations obtained in this way are presented in Table 2.

Stock Company	EDF $\bar{v}_s$	OEDF $\vec{V}_s$	SPT H_s
ACP	0.9709	$\overrightarrow{\mathcal{T}r}(0.9751; 0.9751; 0.9666; 0.9662)$	0.9691
CPS	0.9657	$\overrightarrow{\mathcal{T}r}(0.9699; 0.9657; 0.9657; 0.9632)$	0.9681
ENG	0.9852	$\overrightarrow{\mathcal{T}r}(0.9891; 0.9862; 0.9842; 0.9813)$	0.9806
JSW	0.9615	$\overrightarrow{\mathcal{T}r}(0.9584; 0.9615; 0.9615; 0.9642)$	0.9682
KGH	0.9625	$\overrightarrow{\mathcal{T}r}(0.9592; 0.9599; 0.9650; 0.9678)$	0.9576
LTS	0.9569	$\overrightarrow{\mathcal{T}r}(0.9606; 0.9583; 0.9555; 0.9535)$	0.9685
OPL	0.9652	$\overrightarrow{\mathcal{T}r}(0.9520; 0.9536; 0.9768; 0.9768)$	0.9653
PGE	0.9770	$\overrightarrow{\mathcal{T}r}(0.9789; 0.9789; 0.9751; 0.9732)$	0.9770
PKO	0.9597	$\overrightarrow{\mathcal{T}r}(0.9530; 0.9530; 0.9666; 0.9666)$	0.9597

Table 2 Expected discount factors of portfolio π components. Source: Own elaboration.

The replacement of an accurate PV evaluation by its assessment approximated in a more accurate way, reflects the essence of the PV. If now we estimate PV with the use of TrOFN presented in Table 1 then using the Jensen Alpha goes down to the comparison of an imprecise OEDF with the precise SPT. By means of (15) we then estimate the values of recommendation choice function presented in Table 3.

	Recommendation Choice Function				
Stock Company	A^{--}	A^{-}	A^{0}	A^{+}	A^{++}
ACP	0	1	1	1	0
CPS	0	0.43	0.43	1	0.57
ENG	1	1	0	0	0
JSW	0	0.37	0.37	1	0.63
KGH	1	1	0	0	0
LTS	0	0	0	1	1
OPL	0	1	1	1	0
PGE	0	1	1	1	0
PKO	0	1	1	1	0

Table 3 Imprecise recommendations. Source: Own elaboration.

These are ambiguous recommendations. It means that even the use of precise premises does not guarantee obtaining precise recommendations.

7 Conclusion

Using the Jensen Alpha criterion, we get identical recommendations in each case. For these securities, Sharpe's ratio gave different recommendations for the same portfolio [8]. The next stage of research will be to check what recommendations we will get using the Treynor's ratio and the First Safety Criteria.

References

[1] Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 20*, 87–96.

[2] Dubois, D. & Prade, H. (1978). Operations on fuzzy numbers. *Int. J. Syst. Sci., 9*, 613–629.

[3] Jensen, M.C. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business, 42(2)*, 167–247.

[4] Kosiński, W. (2006). On fuzzy number calculus. *Int. J. Appl. Math. Comput. Sci., 16(1)*, 51–57.

[5] Kosiński, W., Prokopowicz, P. & Ślęzak, D. (2002). Fuzzy numbers with algebraic operations: algorithmic approach. In M. Klopotek, S. T. Wierzchoń &M. Michalewicz (Eds).; Proc.IIS'2002 Sopot, Poland Physica Verlag, Heidelberg, 311–320.

[6] Kosiński, W., Prokopowicz, P. & Ślęzak, D. (2003). Ordered fuzzy numbers. *Bull. Pol. Acad. Sci., 51(3)*, 327–339.

[7] Kosiński, W.& Słysz, P. (1993). Fuzzy numbers and their quotient space with algebraic operations. *Bull. Pol. Acad. Sci., 41(3)*, 285–295.

[8] Łyczkowska-Hanćkowiak, A. (2019). Sharpe's Ratio for Oriented Fuzzy Discount Factor. *Mathematics, 7(3), 272*. 1-16.

[9] Łyczkowska-Hanćkowiak, A. & Piasecki, K. (2018). The expected discount factor determined for present value given as ordered fuzzy number, In W. Szkutnik, A. Sączewska-Piotrowska, M. Hadaś-Dyduch & J. Acedański (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*, University of Economics in Katowice Publishing (pp. 69–75).

[10] Łyczkowska-Hanćkowiak, A. & Piasecki, K. (2018). On representation of Japanese candlesticks by ordered fuzzy numbers. In W. Szkutnik, A. Sączewska-Piotrowska, M. Hadaś-Dyduch & J. Acedański (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*, University of Economics in Katowice Publishing, (pp. 61–68).

[11] Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7(1)*, 77-91.

[12] Nison, S. (1991). Japanese Candlestick Charting Techniques, New York Institute of Finance, New York.

[13] Piasecki, K. (2011). Effectiveness of securities with fuzzy probabilistic return. *Operations Research and Decisions, 21(2)*, 65–78.

[14] Piasecki, K. (2012). Basis of financial arithmetic from the viewpoint of the utility theory. *Operations Research and Decisions, 3*, 37–53.

[15] Piasecki, K. (2014). On Imprecise Investment Recommendations, *Studies in Logic, Grammar and Rhetoric. 37(1)*, 179–194.

[16] Piasecki, K. (2016). The Intuitionistic Fuzzy Investment Recommendations. In A. Kocourek, M. Vavrousek (Eds.), *34th International Conference Mathematical Methods in Economics. MME 2016. Conference Proceedings*, Technical University of Liberec, (pp. 681–686).

[17] Piasecki, K. (2018). Revision of the Kosiński's Theory of Ordered Fuzzy Numbers. *Axioms, 7(1)*.

[18] Piasecki, K. The relations "less or equal" and "less than" for ordered fuzzy number, In W. Szkutnik, A. Sączewska-Piotrowska, M. Hadaś-Dyduch & J. Acedański (Eds.), *10th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Summer Edition". Conference Proceedings* (pp. 32–39).

[19] Piasecki K.& Siwek, J. (2018). Two-Asset Portfolio with Triangular Fuzzy Present Values—An Alternative Approach, In T. Choudhry & J. Mizerka (Eds), *Contemporary Trends in Accounting, Finance and Financial Institutions. Proceedings from the International Conference on Accounting, Finance and Financial Institutions (ICAFFI), Poznan 2016*, Springer International Publishing, (pp. 11–26).

[20] Piasecki, K.& Siwek, J. (2018). Two-assets portfolio with trapezoidal fuzzy present values In W. Szkutnik, A. Sączewska-Piotrowska, M. Hadaś-Dyduch & J. Acedański (Eds.), 9*th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Summer Edition". Conference Proceedings*.

[21] Prokopowicz, P. et al, (Eds.) (2017) *Theory and Applications of Ordered Fuzzy Number. Tribute to Professor Witold Kosiński*; Studies in Fuzziness and Soft Computing, *356*; Springer Verlag, Berlin.

Analysis of Production Function of Agricultural Holdings

Marie Šimpachová Pechrová[1], Ondřej Šimpach[2]

Abstract. Agricultural holdings differ in their production amount due to the size, different yields, technology or weather conditions. The aim of the paper is to find appropriate model for modelling the production function that would account for this heterogeneity and to find the main determinants of the agricultural production. There are panel data available (2 268 observations for 2014–2017), so fixed versus random effects model are considered, compared and tested to choose appropriate model. Cobb-Douglas production function was modelled. The production of farm (sales of own products and services) was explained by the consumption of material, capital, number of employees, and acreage.
Results of fixed effects model were not economically meaningful. Hausmann test proved that random effects model is preferred. Increase of consumed material by 1% caused the increase of production by 0.52%, of tangible long-term assets by 1% caused increase in production by 0.17%, increase of number of employees by 1% caused increase in production by 0.01%, and of land by 0.02%. The production of Czech agricultural holdings is influenced the most by the amount of consumed material, while the number of employees and acreage are less significant.

Keywords: agriculture, fixed effects model, panel data, random effects model

JEL Classification: C23, C51, C12
AMS Classification: 62J05, 62F03

1 Introduction

Agricultural holdings differ in their production amount not only due to the size, but also because of different yields caused by technology or weather conditions. The level of used technology and innovations on the farm is a prerequisite of its level of production and competitiveness. Technology transfer between universities and public or private research institutes and producers can be hence seen as a driver of innovative entrepreneurship in agriculture and the agri-food industry. (Carayannis, Rozakis and Grigoroudis [4]).

The weather changes are particularly important in crop production, but also livestock production can experience the fluctuation in production, especially when pasture system is concerned. For example, Perez-Mendez, Roibas and Wall [16] showed that meteorological variables had a significant impact on milk production (it is higher under warm conditions due to improvements in forage production) in Spanish region of Asturias. Observed and unobserved variables influence the production of agricultural holdings. Therefore, mathematic modelling is useful tool that can help decision making and planning of farmers.

For example, in milk production in New Zealand, a multi-stage stochastic program that incorporates weather dynamics was introduced by Dowson et al. [7] in order to help the decision making and planning of the farmers. Duc-Anh et al. [8] used seasonal climate forecasts in the sugar industry in Australia and demonstrated, that the model can facilitate the irrigation planning of sugarcane farming and consequently contribute to the improvement of gross margin of the growers. Nechaeva et al. [14] identified on a case study of Russia's region Nizdhny Novgorod the factors that influence effective operation of agricultural organizations in terms of their decision making about equipment and machinery. "The research identified two major factors providing a competitive market position of manufacturing companies: factor 1 *Technical and technological modernization of production* and factor 2 *Qualitative composition safety*" (Nachaeva et al., 2019). Daxini et al. [6] examined how the psychological factors can influence the decision making of farmers about adoption of nutrient management practices in Ireland by the behavioural model. The results showed that (on a national sample) attitudes, subjective norms (social pressure), perceived behavioural control (ease / difficulty) and perceived resources were significant and positively associated with farmers' intentions.

[1] Institute of Agricultural Economics and Information, Department of Modelling of the Impacts of Agricultural Policy, Mánesova 75, 120 00 Prague 2, simpachova.marie@uzei.cz.
[2] Institute of Agricultural Economics and Information, Independent Department of Operational Analysis and International Cooperation, Mánesova 75, 120 00 Prague 2, simpach.ondrej@uzei.cz.

Certain effect on production can have also agricultural subsidies provided by the European Union, despite that they shall be decoupled from the amount of production (are not provided on unit of production, but rather on hectare of agricultural land). Pechrová [15] examined to what extend EU's agricultural subsidies influence the production based on the production function of the organic farms in the Czech Republic. She found out that according to the expectation the increase of material, capital, labour and land cause the increase of production and Single Area Payment Scheme (SAPS) and other subsidies mildly decrease the production. Hence, their effect is not significant as it shall be.

The aim of the paper is to find appropriate model for modelling the production function of agricultural holdings that would account for this heterogeneity and to find the main determinants of the agricultural production. There are panel data available, so fixed versus random effects model are considered, compared and tested in order to choose appropriate model. The selection between random and fixed effects model was done for example by Hu and McAleer [11] on a panel data of agricultural production in China. Dataset consisted of 30 provinces for the period 1991–1997. They estimated Cobb-Douglas production function using both, fixed and random effects models, and calculated provincial technical inefficiency. Testing revealed that for this application the random effects model was rejected in favour of the fixed effects model. Also, Heshmati [9] compared the two above mentioned models using a sample of rotating panel data set from Swedish crop producers over the period 1976–1988. He showed that the introduction of heteroscedasticity and the integration of sample selection in the production relationship was important. "The impact of a correction for selectivity bias on the results, in terms of input elasticities and returns to scale is found to be significant" (Heshmati [9]).

2 Data and methods

The aim of the paper is to find appropriate model for modelling the production function of agricultural holdings on panel data. The advantage of panel data is that they can control for unobservable heterogeneity among farms (Ye, Xu and Wu [18]). They allow to control for unobservable variables or variables that change over time but not across entities. "A major motivation for using panel data has been the ability to control for possibly correlated, time-invariant heterogeneity without observing it." (Arellano [1]). "Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency. Panel data are better able to study the dynamics of adjustment and are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data." (Baltagi [2]) Some drawbacks might be difficulties with data collection.

Our analysis utilises accounting data from Albertina database collected by Bisnode company from public firms' register. Acreage of land was obtained from Land Parcel Identification System. There were 2 268 observations for 680 companies for years 2014 to 2017. There were on average 3.3 observations per one firm. All calculations were done in econometric SW Stata version 15.1.

A Cobb Douglas type of production function was constructed where explained variable – the production of farm i in thousands of CZK (y_{it}) – was represented by the sales of own products and services in particular year (t). In order to remove the impact of price changes overtime, the production was deflated by the price index of agricultural producers (2010 = 100).

Explanatory variables included material, capital, labour, and land. Material ($x_{1,it}$) consisted of consumed material and energy by i^{th} farm in time t and capital ($x_{2,it}$) of tangible long-term assets of i^{th} farm in time t. Both variables were deflated by the industrial producers' price index (2010 = 100). Labour ($x_{3,it}$) was represented by number of employees. The acreage of farmland (input land – $x_{4,it}$) was obtained from LPIS database.

A Cobb-Douglas form of production function was selected (1)

$$y_{it} = \beta_0 x_{1,it}^{\beta_1} x_{2,it}^{\beta_2} x_{3,it}^{\beta_3} x_{4,it}^{\beta_4} \tag{1}$$

where β_0 is constant, $\beta_{1,\,2,\,\ldots 4}$ are coefficients of explanatory variables and estimated in linearized form (2):

$$\ln y_{it} = \ln \beta_0 + \beta_1 \ln x_{1,it} + \beta_2 \ln x_{2,it} + \beta_3 \ln x_{3,it} + \beta_4 \ln x_{4,it} + u_{it} \tag{2}$$

If the panel nature of the data is ignored, all data can be pooled together, and the model can be estimated by simple ordinary least square method (OLS). Therefore, firstly a Breusch-Pagan Lagrange multiplier test was used to decide between a random effects regression and a simple OLS regression. The null hypothesis (H_0) is that variances across farms are zero, there is no significant difference across farms (i.e. no panel effect). If the null

hypothesis is rejected, there is random effect present in a model. Both, fixed and random effects models were estimated and compared based on Hausmann test.

Fixed effect model (FEM) requires that all variables change over time otherwise it removes the effect of time-invariant characteristics. Each farm has its own intercept because each farm has its own individual characteristics that may influence the explained variable such as the management, sowing method etc., while the slope coefficient is the same for all. It is assumed that some feature of individual farm may impact or bias the predictor or outcome variables. It is assumed that there is a correlation between farm's error term and predictor variables, but time-invariant characteristics unique to the farm are uncorrelated among themselves. If the error terms are correlated, then FEM is no suitable since inferences may not be correct.

In **random effects model** (REM) the variation across entities is assumed to be random and uncorrelated with the predictor or independent variables in random effects model. It is assumed that differences across farms have an influence on dependent variable. This allows for time-invariant variables to play a role as explanatory variables.

Consequently, **Hausman test** was done to decide between fixed or random effects. It tests whether the unique errors are correlated with the regressors. The null hypothesis is that they are not (difference in coefficients is not systematic), which means that preferred model is random effects vs. the alternative fixed effects. If value W of the Hausman test which follows χ^2 distribution is significant (p-value < 0.05), then random effects model shall not be used. If the p-value is higher than $\lambda = 0.05$, H_0 is rejected and random effect model shall be used.

The **goodness of fit** was assessed by coefficient of multinomial determination. It is calculated as $R^2 = 1 - (\sigma_{ui}^2/\sigma_{yi}^2)$, where σ_{ui}^2 is residual variance and σ_{yi}^2 is total variance in a model.[3]

Statistical **significance of the parameters** was tested by t-test with H_0: the coefficient is not statistically significantly different from 0. Test criterion is calculated as the ratio of the parameter's absolute value (β_k) to the standard error (S_{bi}). If the p-value is lower than $\lambda = 0.05$, H_0 is rejected.

Coefficient of intraclass correlation (ρ) tells that how much of the total variance is given by the differences across panel. It is calculated as the ratio $\sigma_{ui}^2/(\sigma_{ui}^2 + \sigma_{ei}^2)$, where σ_{ui}^2 is variance of residuals within groups and σ_{ei}^2 variance of overall error term.

3 Results

First, FEM was estimated using Stata command `xtreg` and `areg`. Results are displayed in Table 1. Variable land was omitted due to collinearity (it does not change overtime). We considered fixed-effects (within) regression model where there are n entity (firm) – specific intercepts. Correlation between error u_i and the regressors ($\mathbf{X}_b$) was assumed and confirmed in the results (regression coefficient is 0.74). The model is statistically significant as p-value of F test ($F^{[3, 1585]} = 102.15$) indicates that the null hypothesis (H_0: all coefficients in the model are jointly equal to zero) is rejected. Also p-value of F-test for testing that all within groups residues u_i are equal to 0 (value $F^{[679, 1585]} = 23.93$) showed that H_0 is rejected. At least one error is different from zero. According to ρ 95.40% of variance is due to differences across panels.

The value of the coefficients does not correspond to the expectations. Two-tail p-values of t-test for each coefficient shows that only constant and consumption of material and energy are statistically significant. If it increases by 1%, then the production increases by 0.34%. Capital is insignificant determinant of the production as same as the number of employees. The second mentioned coefficient is even negative, indicating possible overemployment when the increase of the number of employees cause decrease of production. Hence, the model itself is not optimal.

Another way of estimation of fixed effect model was done to compute real R^2 value. It has again n entity-specific intercepts and the results are similar. R^2 was 98.22% and after adjustment 97.46% which indicates high goodness of fit of the model. This means that the variability of explained variable is explained from 97.46% by variability of explanatory variables.

Because the results of fixed effects model were not satisfying, random effects model was estimated too. First, it was tested whether to use random effects model or simple OLS regression by Breusch and Pagan Lagrangian multiplier test. χ^2 value of the test was 1476.77 with p-value 0.00. Hence, we rejected H_0. There are significant differences among the farms and random effect model is appropriate. Simple OLS regression cannot be done.

[3] The command `xtreg` in SW Stata 15.1 provides meaningless coefficient of multinomial determination (R^2) and adjusted R^2. Therefore, command `areg` was utilized to report correct values of R^2.

Variable	Coef.	Std. error	t-value	p-value
Constant	6.8437	0.2715	25.2100	0.0000
ln x1	0.3415	0.0199	17.1600	0.0000
ln x2	0.0183	0.0225	0.8100	0.4160
ln x3	-0.0018	0.0012	-1.4900	0.1360
ln x4	(omitted)			
σ_u	1.3642		σ_e	0.2996
ρ	0.9540		corr (u_i, $\mathbf{X}_b$)	0.7359
$F^{[3, 1585]}$	102.1500		p-value (F)	0.0000
$F^{[679, 1585]}$	23.9300		p-value (F)	0.0000
R^2	0.9822		R^2 adjusted	0.9746

Table 1 Results of fixed effects model

Therefore, random effects model was estimated. Results are displayed in Table 2. The model as a whole was statistically significant as the probability of Wald χ^2 = 2703.09 was 0.00. All estimated coefficients were statistically significant too, but again with the exception of the number of employees. We can conclude that the number of employees does not statistically significantly influence the amount of production of agricultural companies in a sample as both models considered it statistically insignificant. Maybe due to seasonality of the work in agriculture, the labour shall be measured in Annual Work Units (AWU) to be statistically significant determinant but our dataset does not contain it. Besides, agriculture is labour demanding mainly in case of livestock production. Crop production requires less employees and (mostly) seasonal. The sample is mixed but might be with the predominance of crop production. Values of other coefficients corresponds to the expectations and are statistically significant. Increase of consumption of material and energy by 1% brings increase of production by 0.52%, the elasticity is the highest here. Then the increase of capital by 1% brings increase of production by 0.17%. Increase of acreage of land by 1% brings increase of production by 0.02%. The lowest elasticity is in case of number of employees, but it is not a significant determinant.

Variable	Coef.	Std. error	t-value	p-value
Constant	3.5058	0.1495	23.4600	0.0000
ln x1	0.5217	0.0150	34.8500	0.0000
ln x2	0.1704	0.0167	10.2000	0.0000
ln x3	0.0008	0.0013	0.6600	0.5110
ln x4	0.0193	0.0057	3.4100	0.0010
σ_u	0.8693		σ_e	0.2996
ρ	0.8938		corr (u_i, $\mathbf{X}_b$)	0 (assumed)
Wald $\chi^{2[4]}$	2703.09		p-value (F)	0.0000
R^2	0.7948			

Table 2 Results of random effects model

Hausman tests ($\chi^2_{[3]} = (b-B)^T \left[V_b - V_B\right]^{-1} (b-B) =$ 216.87, p-value = 0.00) showed that null hypothesis is rejected, the difference in coefficients are not systematic, and fixed effects can be used. Farms' errors (u_i) are not correlated with the regressors, which is the assumption of fixed effects model. Hence, random effect model is more appropriate to model the relation between production and production factors. This model also complies with economic assumptions set prior the estimation. Also, random effects model can deal with regressors that are fixed across individual farms (land in our case). Random effects models offer distinct advantages over fixed effects models in terms of their efficiency and their ability to calculate shrunken residuals (Clarke et al. [5]).

Also, the total residual variance can be partitioned into two components: the between-farm variance σ_u^2 and the within-farm variance σ_e^2. In our case, the variance between farms in higher (0.87) than within farm (0.30). According to ρ 89.38% of variance is due to differences across panels.

On the other hand, there are certain drawbacks. There is a certain probability that there is a correlation between the unobserved effects and the explanatory variables. In fixed effects model it is assumed that there is certain

correlation, but in random effects model is considered to be zero, that does not have to be true. This implies inconsistency due to omitted variables. On the other hand, fixed effects model is consistent despite being inefficient.

The correlation problem can be overcome by fitting correlated random-effects and correlated random-coefficient models. It is used for example by Suri [17] on the decision making of Kenyan farmers to adopt new technologies where the benefit and costs of technologies are heterogenous, so the farmers with low net returns do not adopt new technology. It his case, the returns to new technology (hybrid wheat) are heterogeneous and correlated with the decision to use hybrid, so he estimates correlated random coefficient model. For an ease of the calculation, there is a community-contributed command `randcoef` in Stata programmed and explained by Cabanillas, Michler and Michuda et al. [3].

Unobserved heterogeneity in panel data modelling concerned Hsiao [10]. He used various formulations of the interactive effects models and suggested a quasi-likelihood approach as a framework to study issues of estimation and statistical inference when regressors are either strictly exogenous or predetermined and under different combinations of the data size. Also, Musafiri and Mirzabaev [13] noted that the results of their cross-section model shall be treated with caution due to unobserved heterogeneity among agricultural households in Rwanda which they surveyed. Therefore, they constructed also panel data model – particularly pooled OLS (difference in difference) model and fixed effects model to observe the changes in agricultural output over time. "The results confirm the predominant role of labour, capital, and land quality to output growth over time." (Musafiri and Mirzabaev [13]).

It is therefore reasonable to assume that production factors and other differences across farms have an influence on production because each farm has slightly different structure of production, is located in different weather conditions and is managed variously. All those unobserved characteristics can affect the volume of production. For example, it was proved by Perez-Mendez, Roibas and Wall [16] that meteorological variables can affect milk production through their impact on the productivity of cows and the production of foodstuff. They proposed a production function where meteorological variables affect the productivity of cows and the production of forage, thereby indirectly affecting milk production.

Mundlak and Butzer [12] analysed the role of capital in agricultural production in the cross-country study of production functions. The results of their model revealed the relative importance of capital, because the lack of it can be a constraint on agricultural growth. In our case, the highest intensity on production had the consumption of material and energy. The coefficient (which is also an elasticity coefficient) shows that 1% change of the consumption causes 0.52% change in production. Capital and land changes are less important. The lowest elasticity had the labour which was even statistically insignificant. It can be due to the fact that "The shift to more productive techniques is associated with a decline in labor, reflecting labor-saving technical changes" (Mundlak and Butzer [12]).

Conclusion

Agricultural holdings differ in their production amount not only due to the size, but also because of different type of production, yields caused by technology or weather conditions. The aim of the paper was to find appropriate model for modelling the production function of agricultural holdings that would account for this heterogeneity and to find the main determinants of the agricultural production. Fixed versus random effects model were considered, compared and tested in order to choose appropriate model. A sample of agricultural holdings consisted of 2 268 observations for 680 groups of farms for years 2014–2017. Production function was modelled in Cobb-Douglas form that has an advantage that the coefficients can be interpreted as elasticities. Production was explained by material, capital, labour, and land. Accounting data were taken from Albertina database. The acreage of farmland was obtained from LPIS land register.

Then we tried to decide whether to use fixed effects or random effects model to estimate the production function of agricultural firms. The nature of the data would suggest using random effects model. It is due to the fact, that the acreage of land does not change during the observed period of time. For this reason, this variable will be omitted from fixed effects model. However, fixed effects model still might be better to estimate the production function of agricultural holdings.

Results provided by fixed effects model were not economically feasible. Coefficient in case of labour was negative suggesting that increase of number of employees would bring decrease in production. Probably AWU would be more suitable variable, but it was not available for us. Also, as the acreage has not changed much during years, it was omitted from the calculation.

Hausmann test also proved that random effects model is preferred to fixed effects. Therefore, a random effects model was estimated. The results corresponded to the economic theory that increase of consumed material by 1% cause the increase of production by 0.52%, increase of usage of tangible long-term assets by 1% cause the increase of production by 0.17%, increase of number of employees by 1% increase the production by 0.01% and increase of land by 1% cause increase by 0.02% in production. The production of Czech agricultural holdings is influenced the most by the amount of consumed material, while the number of employees and acreage play minor role (there were 11% of farms with only livestock production).

Acknowledgements

The paper was supported by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0918, Internal Research Project no. 1113/2019.

References

[1] Arellano, M. (2003). *Panel Data Econometrics*. New York: Oxford University Press. ISBN 0-19-924528-2.
[2] Baltagi, B. H. (2005). *Econometric analysis of Panel Data.* New York: John Wiley & Sons Inc. ISBN 0-470-01456-3.
[3] Cabanillas, O. B., Michler, J. D. & Michuda, A. et al. (2018). Fitting and interpreting correlated random-coefficient models using Stata. *Stata Journal*, ***18***(1), 159–173.
[4] Carayannis, E. G., Rozakis, S. & Grigoroudis, E. (2018). Agri-science to agri-business: the technology transfer dimension. *Journal of Technology Transfer*, ***43***(4), 837–843. DOI: 10.1007/s10961-016-9527-y.
[5] Clarke, P., Crawford, C., Steele, F. & Vignole, A. (2010). The Choice Between Fixed and Random Effects Models: Some Considerations for Educational Research. *Discussion Paper No. 5287.* IZA DP No. 5287. Available from http://ftp.iza.org/dp5287.pdf.
[6] Daxini, A., O´Donoghue, C., Ryan, M., Buckley, C., Barnes, A. P. & Daly, K. (2018). Which factors influence farmers' intentions to adopt nutrient management planning? *Journal of Environmental Management, **10***, 350–360.
[7] Dowson, O., Philpott, A., Mason, A. & Downward, A. (2019). A multi-stage stochastic optimization model of a pastoral dairy farm. *European Journal of Operational Research*, ***274***(3), pp. 1077–1089. DOI: 10.1016/j.ejor.2018.10.033.
[8] Duc-Anh, A., Mushtaq, S., Reardon-Smith, K., Kouadio, L., Attard, S. & Cobon, D. (2019). Value of seasonal forecasting for sugarcane farm irrigation planning, *European Journal of Agronomy*, ***104***, 37–48. DOI: 10.1016/j.eja.2019.01.005.
[9] Heshmati, A. (1994). Estimating random effects production function models with selectivity bias: an application to Swedish crop producers. *Agricultural Economics, **11***(2–3), 171–189. DOI: 10.1016/0169-5150(50)00017-3.
[10] Hsiao, Ch. (2018) Panel models with interactive effects. *Journal of Econometrics*, ***206***(2), 645–673.
[11] Hu, B. & McAleer, M. (2005). Estimation of Chinese Agricultural Production Efficiencies with Panel Data. *Mathematics and Computers in Simulation*, ***68***(5–6), 474–483.
[12] Mundlak, Y. & Butzer, R. (1997). The Determinants of Agricultural Production: A Cross-Country Analysis. *Policy Research Working papers.* The World Bank. https://doi.org/10.1596/1813-9450-1827
[13] Musafiri, I. & Mirzabaev, A. (2014). Long-Term Determinants of Agricultural Output in Smallholder Farmers in Rwanda. In *Proceedings of Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting* (pp. 1–39). Minneapolis: Agricultural & Applied Economics Association.
[14] Nechaeva, M. L., Nechaev, V. N., Ilicheva, O. V., Kozlova, L. A. & Kurilova, A. A. (2019). Material and Technical Base of Agricultural Organizations as the Basis for their Economic Efficiency. *International Journal of Advanced Biotechnology and Research*, ***10***(1), 363–370.
[15] Pechrová, M. (2014). The Impacts of the EU's Subsidies on the Production of Organic Farms. In Jedlička, P. (Ed.), *Proceedings of Hradec Economic Days.* Hradec Králové: University of Hradec Králové, 124–130. ISSN 2464-6067.
[16] Perez-Mendez, J. A., Roibas, D. & Wall, A. (2019). The influence of weather conditions on dairy production. *Agricultural Economics*, ***50***(2), 165–175. DOI: 10.1111/agec.12474.
[17] Suri, T. (2011). Selection and Comparative Advantage in Technology Adoption. *Econometrica*, ***79***(1), 159–209.
[18] Ye, X., Xu, J. & Wu, X. (2018). Estimation of an unbalanced panel data Tobit model with interactive effects. *Journal of choice modelling*, ***28***(C), 108–123.

Mixtures of Probability Distributions

Luboš Marek[1], Michal Vrabec[2]

Abstract. Probability distributions are commonly used for modelling data governed by interval frequency distributions. The exact shape of a typical distribution of this kind is determined by the parameters of that distribution. In other words, our task is reduced to parameter estimation and, based on that, it is no problem to construct the exact shape of the distribution. A suitable model can thus be found for existing data, whether past or present. We show in our article that significantly better results can be achieved by using a mixture of probability distributions. This mixture can be formed in one of two ways. The first way is based on the unknown relevance of the observation to the component in the mixture, and the second way is based on the known relevance of the observation to the component in the mixture. We will use the latter and show the application of a mixture distribution with a different number of components in the mixture to the wage data in the Czech and Slovak Republics. Using time series analysis, we will also propose a way to predict future wage distributions in both countries.

Keywords: wage, wage characteristics, probability distribution, mixture of distribution, parameter estimates.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In this paper, we analyse the time evolution of wage distributions in the Czech and Slovak Republics in the most recent 18 years (namely, 2000 through 2017). We cannot yet gain access to the data for the year 2018 at the time of writing this paper. Several works have been analysing and describing the wage distributions – cf. Marek, Vrabec [5], and Vrabec, Marek [9]. The standard course of such an analysis is to use observations from the past to create our model, and the shape of the distribution is derived from the past data. Each continuous distribution of probability has the corresponding density of probability, and we choose the distribution to maximise the fit of the theoretical curve to the empirical data. The distributions are parametric; that is, the shape of the density depends on the values of parameters, which must be estimated from the past observations. After that we can draw the density curve as a model suitable for the wage distribution. This is the usual approach and the most frequently used density types include the normal, lognormal, logistic, and some other distributions.

We want to show below that a far better approach is to make use of probability distribution mixtures instead of individual distributions.

Another problem is to draw the shape of the future distribution of wages. Here we use estimates of the distribution parameters based on the data for all 18 years, obtaining a time series of the parameters, from which we, with the aid of time series analysis, derive estimates of the parameters to be used in predicting the future shape of the wage distribution.

We tested this principle on mixtures of normal and lognormal distributions, and the results achieved appear to be very good. The actual creation of the mixtures may follow the algorithms described in the literature – cf. Malá [4], and Bartošová, Longford [2] – in particular, regarding the numbers of curves in individual mixtures. On the other hand, mixtures are created in a natural way, based on the data division by different factors. The wage data can, for example, be categorised by gender, age, education, region, etc. Such categories arise naturally, creating logical mixtures represented by individual densities (and their parameters). At this point we are ready to utilise these densities for creating meaningful mixtures.

[1] University of Economics, Prague, Department of Statistics and Probability, W. Churchill sq.4, 130, 57 Prague 3, marek@vse.cz.
[2] University of Economics, Prague, Department of Statistics and Probability, W. Churchill sq.4, 130, 57 Prague 3, vrabec@vse.cz.

2 Data and Results

We work with datasets containing the wage data from the Czech and Slovak Republics in the years 2000 through 2017, i.e., an 18-year time series of average wages. In addition to the average values, certain additional characteristics have been studied, namely, variability and important quantiles. For each of the studied factors we have a table of the interval distribution of frequencies (with the width of each interval at 20 EUR). We can derive the empirical distributions of wages from these tables. Our data come from the Czech and Slovak Trexima Companies [7], [8]. In order to compare wage data from both countries we must convert all data to EUR and adjust the data to inflation, taking the exchange rate from the national banks' websites and the inflation rate values from the national statistical offices' websites. The entire analysis has been carried out in MS Excel (simple tables and charts) and in the JMP software (calculations on density mixtures and parameter estimates). The entire process of setting up the mixtures is demanding from both theoretical and computational viewpoints.

2.1 Evolution of Wage Distributions

First, we will have a look at the empirical wage distributions in both countries – cf. Figs. 1 and 2.

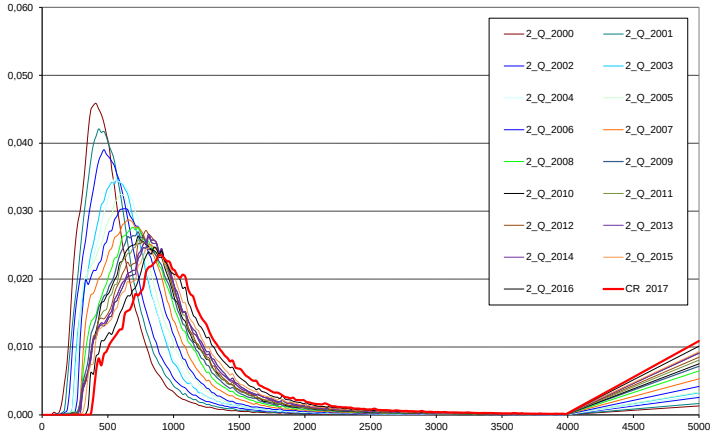


Figure 1 Empirical wage distribution – ČR

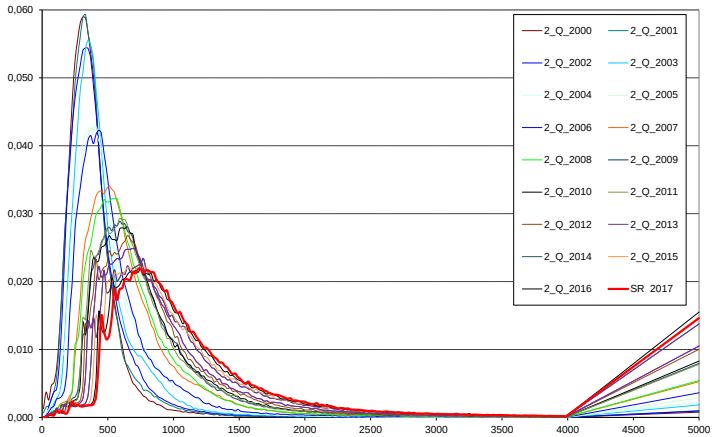


Figure 2 Empirical wage distribution – SR

Both charts are drawn at the same scale to emphasise the differences between the countries. Our goal is to find the model for the distributions up to 2017 and estimate the evolution in 2018.

Finally, let us point out that we have only modelled the wage distribution for wages up to a limit of 4,000 EUR (per month). Only very few employees' wages get over that limit. The importance and analysis of high wages (defined as those higher than 4,000 EUR per month in this paper) were analysed in detail in Marek [6].

2.2 Model

First, we identify a model for the empirical distribution of wages with the aid of a sole distribution of probability. We have gone through several density types when searching for a suitable distribution. The model based on the log-logistic density has been identified as the best one. This conclusion confirms our previous results – cf. Marek, Vrabec [5], where this distribution is described in more detail.

However, using density mixtures has proved to be the most useful approach. We have tried using different numbers of curves in the mixture (according to individual factors by which the data is categorised). As a rule, the more components in the mixture the better the outcome. In our case, the best is a mixture of normal distributions containing 14 components corresponding to the regions in the Czech Republic and eight components corresponding to the regions in the Slovak Republic. Each normal distribution has two parameters; hence we get 2 × 14 parameters in the Czech Republic in each year, and 2 × 8 parameters in the Slovak Republic each year. In other words, the entire process is demanding both theoretically and numerically. When deriving the mixtures of normal distributions, we must determine the parameters of the resulting mixture as well as the final shape of the density to be estimated. For *k* components, the latter equals

$$f(x/\mu,\sigma,\pi)=\sum_{i=1}^{k}\frac{\pi_i}{\sigma_i}\phi\left(\frac{x_i-\mu_i}{\sigma_i}\right) \tag{1}$$

where μ_i,σ_i,π_i are: the expectation, standard deviation, and the weight of the *i*-th distribution in the mixture; $\phi(.)$ is the probability density of the standard normal distribution. The parameters of the mixture components are estimated from the past observations with the aid of time series analysis. The expectation

$$E(X)=\sum_{i=1}^{k}\pi_i\mu_i \tag{2}$$

and the variance

$$Var(X)=\sum_{i=1}^{k}\pi_i(\mu_i^2+\sigma_i^2)-\left(\sum_{i=1}^{k}\pi_i\mu_i\right)^2 \tag{3}$$

are the parameters of the resulting mixture. We will present the model for the 2017 mixture and compare it with reality. The data from the years 2000 through 2016 has been used for deriving the model, and the year 2017 has been used as a reference year. The data from the year 2018 is not yet known at the time of writing this paper – hence we estimate the future shape of the distribution in 2018, and all the data from the years 2000 through 2017 has been used when deriving this model. As soon as the data from 2018 is known, we can compare our prediction with the actual data and evaluate the quality of our model. Now we display our results for the Czech Republic.

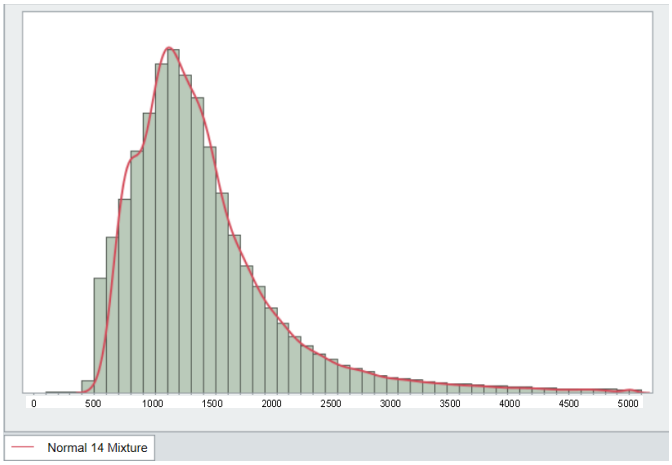


Figure 3 Mixture of normal distributions – CR 2017

The red line shows the resulting curve of the mixture of normal distributions, the columns show the empirical frequencies. This Figure illustrates the fact that the quality of our model is very good. Our calculations are related to many aspects of the outcome; in particular, these aspects include: the estimated values of the parameters, the calculated weights of individual components in the mixture, the quantiles, and the AIC and BIC model quality criteria – cf. Akaike [1] and Chen, Chen [3] for more details. We do not present these values here due to space considerations – just for the data related to the Czech Republic they include $3 \times 14 = 42$ parameter values.

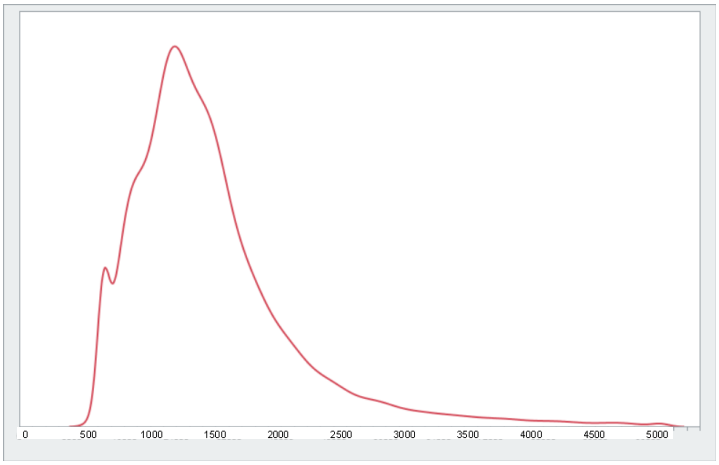


Figure 4 Mixture of normal distributions – prediction of CR 2018

Now we will have a look at the results for the Slovak Republic. The 2017 model, derived in a way similar to the Czech Republic, is shown in Figure 5 (i.e., this Figure shows a comparison of the model with the empirical data valid for the year 2017).

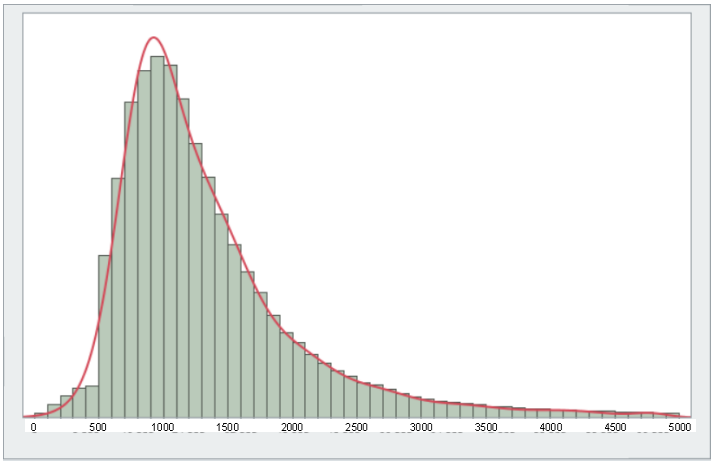


Figure 5 Mixture of normal distributions – SR 2017

The last Figure shows the predicted shape of the distribution for 2018. Again, the data from all the observation years have been used for deriving this model, and no comparison with the 2018 actual data is possible because this data for the Slovak Republic is not yet available at the time of writing this paper.

In none of the instances considered is the predicted shape of the density a smooth curve. We have to keep in mind that we are estimating density mixtures from empirical data and, therefore, the resulting curve cannot be as smooth as for single distributions.

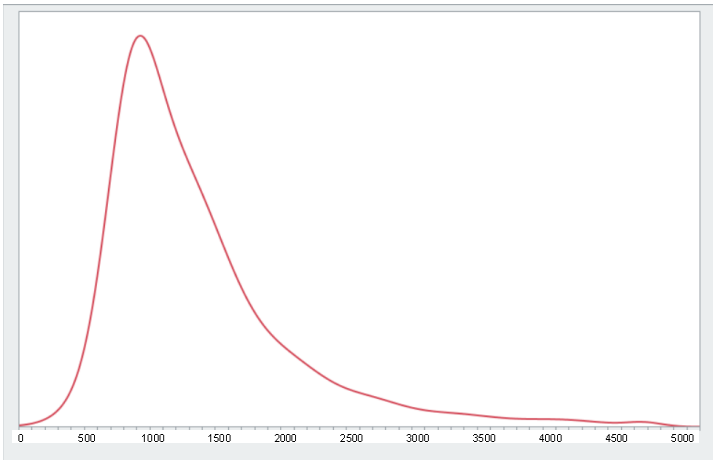


Figure 6 Mixture of normal distributions – prediction of SR 2018

Conclusions

The goal of our analysis is to find the optimum approach to modelling wage distributions in the Czech and Slovak Republic. We have tested different models, based on both a sole probability distribution and a mixture of probability densities. For the former, the log-logistic distribution has been identified as the best option. However, the latter approach has turned out to be better – namely, for mixtures of normal distributions. They are admittedly much more demanding from the theoretical and computational viewpoints, but they provide better results. As a rule, the higher the number of components in a mixture the better the resulting model. The best quality has been achieved by a mixture of normal distributions containing 14 components (corresponding to the 14 Czech regions) in the Czech Republic and eight components (corresponding to the eight Slovak regions)

Comparing the models derived for the CR and SR, the outcome is, with respect to the probability distributions used, the same for both countries. The reason may be that there are no significant differences between the empirical distributions of wages in the CR and SR. The wage situation is about the same in both countries, especially in the most recent years

Acknowledgements

This paper was written with the support of the Institutional Support to Long-Term Conceptual Development of Research Organisation, the Faculty of Informatics and Statistics of the University of Economics, Prague.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, vol. 19 , issue 6.
[2] Bartošová, J. – Longford, N. T. (2014). A study of income stability in the Czech Republic by finite mixtures, Prague Economic Papers, vol. 23, no. 3, pp. 330–348..
[3] Chen, J.; Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. Biometrika. 95 (3): 759–771. CiteSeerX 10.1.1.505.2456. doi:10.1093/biomet/asn034.
[4] Malá, I. (2017) L-moments of a finite mixture of probability distributions. In: International Days of Statistics and Economics (MSED) [online]. Praha, 14.09.2017 – 16.09.2017. Slaný: Melandrium, Libuše Macáková, 2017, s. 929–938. ISBN 978-80-87990-12-4.
[5] Marek, L. – Vrabec, M. (2013). Model wage distribution – mixture density functions. International Journal of Economics and Statistics [online]. 2013, roč. 1, č. 3, s. 113–121. ISSN 2309-0685.

[6] Marek, L. (2017). The effect of extreme wages on average wage values. In: International Days of Statistics and Economics (MSED) [online]. Praha, 14.09.2017 – 16.09.2017. Slaný: Melandrium, Libuše Macáková, 2017, s. 962–969. ISBN 978-80-87990-12-4.
[7] Trexima.cz (2019): http://www.trexima.cz
[8] Trexima.sk (2019): http://www.trexima.sk
[9] Vrabec, M. – Marek, L. (2016). Model for distribution of wages. In: AMSE 2016 (Applications of Mathematics and Statistics in Economics) [online]. Banská Štiavnica, 31.08.2016 – 04.09.2016. Banská Bystrica: Občianske združenie Financ, 2016, s. 378–386. ISBN 978-80-89438-04-4. ISSN 2453-9902.

The Dynamics of Human Agent in Project Management

Dominik Škoda[1], Igor Krejčí[2]

Abstract. The paper deals with the application of system dynamics to explain the principles of work effort development in project management. The paper follows the previous research in the field that focused on the modelling of the work effort, however, incorporated functions that lack the explanatory function, do not pass the dimensional analysis and work mainly as a black box. Among others understanding the project as a dynamic complex system means understanding its feedback structure of interconnections between its elements and characteristics, the nonlinearity of cause and effect and delays.
The presented model comprehends the modelling structures that represent effects of initial motivation, understanding and the adaptation of the perceived project difficulty and time demandingness parameters, procrastination and fatigue. As a result, phenomena such as student syndrome and Parkinson's law described in the project management literature are possible to simulate and interpret.
The last part of the paper shows the basic simulation scenarios of behaviour and work effort development that depicts the possibility of parameter settings and calibration of the management flight simulator.

Keywords: Computer simulation, Project management, Scenarios, System dynamics, Work effort.

JEL Classification: C44, C63
AMS Classification: 90B99, 93C15

1 Introduction

Due to the impact on project duration, earned value management and budgeted cost for work performed [2], [3] the work effort of an allocated resource in project management has been deeply discussed and mathematically formulated [1], [16]. This paper follows the research of [4] and [12] that stress the necessity of work effort modelling including such phenomena as Parkinson's Law [18] and Student's syndrome [9]. That kind of work effort development is variable over time in time and single mathematical function for its representation must have more than one local extreme. Therefore, authors apply the trigonometric functions [4], [11] or 4^{th} order polynomial [2], [3].

System dynamics approach to project management is based on a holistic point of view, which take into consideration the feedback structure of the complex project [10], [24]. System dynamics model of the project provides the tool to describe, explore and interpret the behaviour within the project management three major areas – monitoring and control, rework generation and human resource management [22]. However, understanding the project as a system (i.e. application of system thinking on the project management) requires the system thinking skills [21]. Among these skills, structural thinking means thinking in terms of units of measure. The trigonometric functions and polynomials for the work effort development can't satisfy the dimensional consistency test [5]. When we describe the project system in terms of the system of equations, the units of measurement of every variable in every equation must be consistent and without the necessity of artificial parameters that don't have the real-world background. Moreover, when the goal of the modelling is the understanding and interpretation of the behaviour, the functions describing the effect of one variable on another should be non-increasing or non-decreasing, otherwise, it comprehends more effects, which should be described separately [23]. From this perspective, both goniometric functions and higher order polynomials for the description of work effort serves as a black box and limits the explanation function of the model.

In our paper, we present the system dynamics simulation model that describes the work effort in the project management process. Subject to the initial settings, the model allows simulating various situations and behaviour

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems of Engineering, Kamýcká 129, Prague 6 - Suchdol, Czech Republic, xskod002@studenti.czu.cz.
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems of Engineering, Kamýcká 129, Prague 6 - Suchdol, Czech Republic, krejcii@pef.czu.cz.

of the source. In the last part of the paper, we present basic simulation scenarios that describe various settings and important phenomena of work effort development.

2 Material and Methods

The research followed the basic process of system dynamic modelling [8], [15]. The first step of the modelling process resulted in dynamic hypothesis and was formalised as a causal loop diagram. Figure 1 contains the simplified stock and flow diagram of the project resource development. The stock and flow diagram has a close connection to the mathematical formulation, in this case, the computer simulation that is based on the system of first-order differential equations [25].

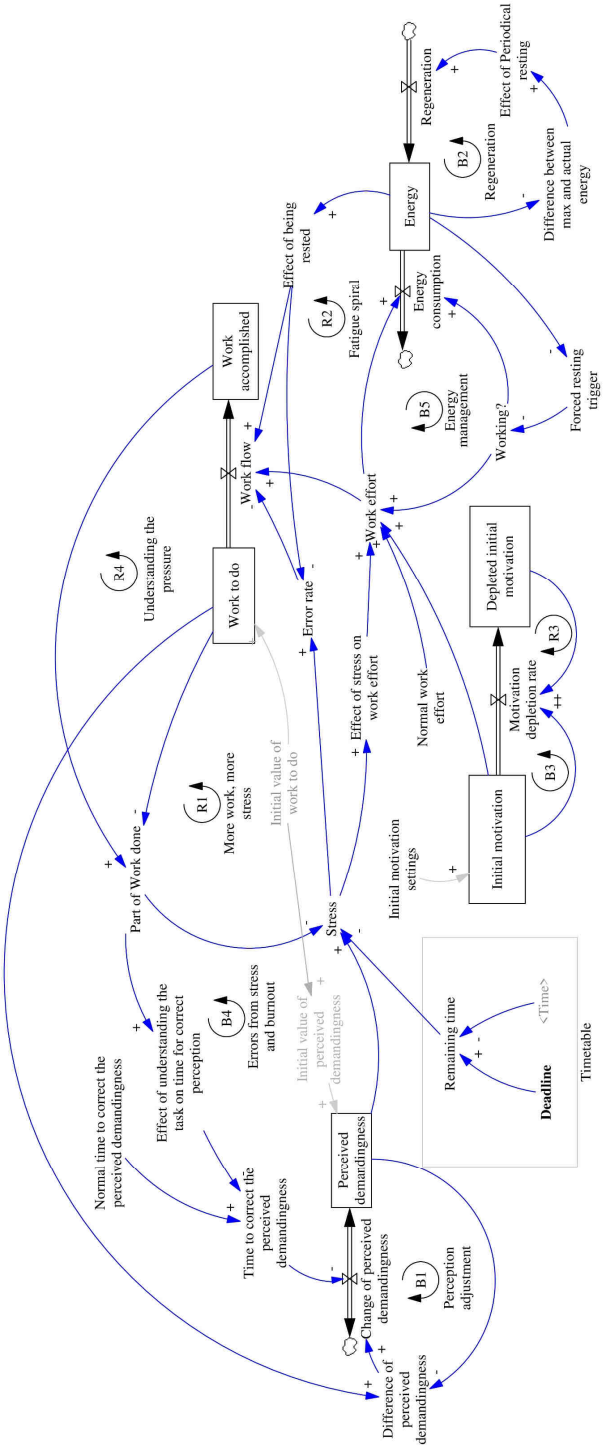


Figure 1 Stock and flow diagram of human agent in project management

Boxes in figure 1 represent the stock variables (levels, accumulations), flow variables are denoted by the pipes. The diagram contains the causal links of two possible polarities with clear mathematical interpretation. The positive link is interpreted as [23]:

$$\frac{\delta y}{\delta x} > 0, \tag{1}$$

the negative link polarity has the opposite interpretation [23]:

$$\frac{\delta y}{\delta x} < 0. \tag{2}$$

Stocks *s* are represented in a mathematical model as the definite integrals [23]:

$$s = \int_{T_0}^{T} (i - o)\mathrm{d}t + s_{T_0}, \tag{3}$$

where *i* are all inflows, *o* are all outflows, T_0 is initial time, *T* is current time and *t* is any time between *T* and T_0.

Figure 1 also shows the most important feedback loops that determine the dynamics of the system. Positive, self-reinforcing loops are denoted as *R*. According to the equations (1) – (3), the loop of causal links in such feedback means that the change in the variable is even strengthened because of the interrelations of the variables. On the contrary, the balancing (negative) loop *B* counteracts the change [14].

Since the value of stock variables could be changed only by inflows and outflows, there shouldn't be the causal link from the parameter or auxiliary variable pointing on the box variable. The exception is the setting of the initial value of the variable. For these purposes, we used the grey arrows and grey font for the initial value.

The very basic structure "Work to do" -> "Work accomplished" is a common part of the system dynamics models of project management (e.g. [7], [13] and [17]), however, usually used for the rework and backlog structure. Our model focuses mainly on the human agent influence similarly to [3], [4]. As the project proceeds, the human resource is more accurate with the estimation of the project demandingness, the difference between remaining time and the perceived demandingness could result in the stress that leads to the increased work effort and strengthens the effect of core reinforcing feedback loop. The increased work effort depletes the energy and causes an increase of the errors that decrease the work flow. The fatigue spiral counteracts the previous loop when the increased pressure decreases the impact of work effort because of burnout. The simulation model comprehends the settings of the level of initial motivation and the type of effect of the stress on the work effort (the different maximal effort and the different threshold of the stress to increase the work effort).

Table 1 provides the model boundary – the most significant endogenous and excluded variables [19], the presented model doesn't contain any exogenous variable, therefore, this common part of the model boundary table isn't presented.

Endogenous variables	Excluded variables
Work effort	Staff (resource) hiring
Error rate	Staff (resource) turnover
Perceived demandingness	Skills and experience level
Motivation depletion	
Stress effect on the work effort	
Energy	

Table 1 Human agent in project management model – model boundary

3 Results and Discussion

The student syndrome from the project management perspective results into the uneven intensity of work effort where initial effort falls due to the comfort and/or procrastination of the human, consequently the lower work effort results in the time pressure and rise of the work effort [4], [9]. Parkinson's law simply says that the activity takes the whole time devoted to it [18]. Figure 2 shows the three basic scenarios that simulate the task completion

according to this phenomena. With the same initial motivation but the different effect of the stress from time pressure on the work effort and different maximal work effort. In all scenarios the deadline is the 100th day, demandingness of the task is 100 work days of normal work effort and initially perceived demandingness is 60 work days. Since the higher work effort must be more energy consuming the right side of the figure shows the development of the level of energy throughout the whole task completion.

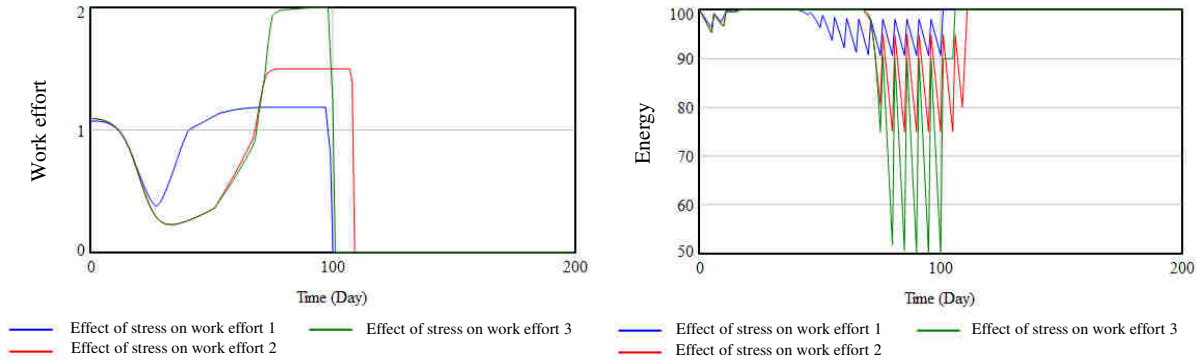


Figure 2 Impact of various stress effects

The initial motivation is depleted, the estimation of the demandingness is improved throughout the initial effort, however, the work effort falls until the time pressure reaches the threshold that activates the stress effect and the reinforcing loop *R*1 become dominant. The effect of stress in first scenario is characterised by low maximal work effort but also by the high sensitivity on the time pressure, therefore the task is finished on time. Second and third scenarios show the same sensitivity but different maximal work effort. Due to the low sensitivity on time pressure and not enough maximal work effort the second scenario simulates late task completion.

Figure 3 shows a different situation when the task of same demandingness as the previous has a different deadline. In this case, it is not necessarily the real deadline, but the shorter deadline could be also used as the buffer – the real deadline is later but the workers assigned to the task are informed about the earlier deadline. This approach is sometimes recommended to deal with the student syndrome and Parkinson's law [4], [9]. The four presented scenarios have the same real and perceived demandingness as the previous situation and high maximal work effort, however, they differ in deadline (60, 100, 160, 200).

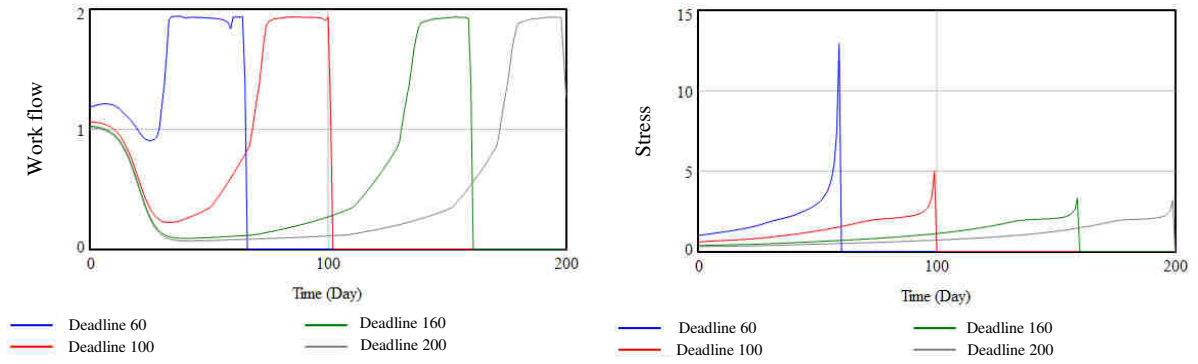


Figure 3 Impact of the deadline

In this case, the left side of figure 3 doesn't show the work effort but the work flow, which is flow between work to do and work done. This variable represents the work effort affected by the fatigue and error rate. The stress on the right is the ratio of the part of work done of perceived demandingness and the remaining time. The shortened deadline results in the small delay (66th day) but the task is finished before the real deadline on the 100th day. On the other hand, when the deadline is prolonged the task (according to Parkinson's law) will not be finished before deadline. The task took all the devoted time and the increased work effort is similar to the standard settings. From the managerial point of view, the question in this simulation is the ethical aspect of the high stress – the size of the buffer must be set accurately to avoid hurting of the workers and also demotivation. The lower maximal work effort would lead to greater delay and possible feeling of the failure of the employees, despite the task is finished before the real deadline.

Figure 4 shows the increased impact of the energy. The previous situations assume doubled work effort for part of the task completion period. However, the fatigue with such effort must grow and the periodical regeneration couldn't be enough. The modelled consequences of fatigue are increasing error rate and decreasing work flow. The energy is consumed when the work effort exceeds the standard level equal to 1. The first scenario works with the standard energy consumption of 10% of total energy per unit over standard level of work effort. Other two scenarios demonstrate the consumption increased to 15% and 20%. The actual settings work with periodical regeneration after 5 work days.

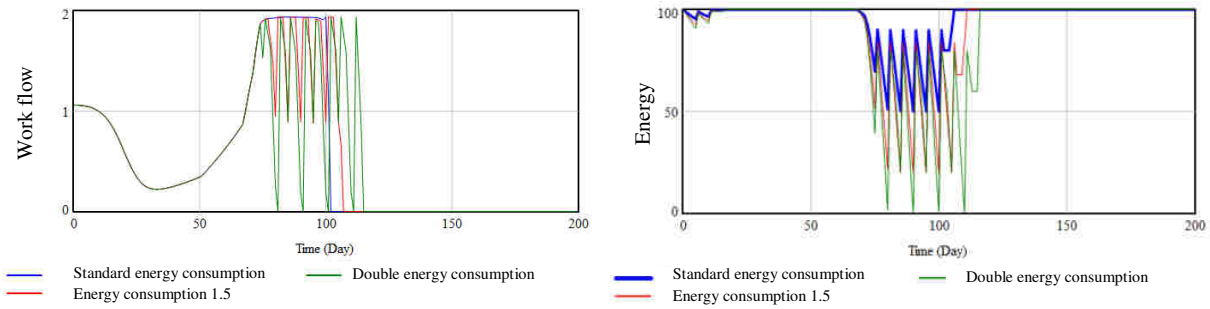


Figure 4 Energy consumption

The both scenarios with increased energy consumption lead to late task completion. The doubled energy consumption even leads to forced resting when the energy is depleted. The increased energy consumption leads to the dominance of the loop that increases the fatigue and counteracts the loops increasing the work flow.

4 Conclusion

The article interprets the effect of the human agent in project management through the system dynamics simulation model. The presented model reacts to the research that interpret the influence of the human agent in project management on the basis of the single equation. The necessity of direction changes led the referred authors to application of trigonometric functions and higher order polynomials. However, such equations lack the interpretation function of individual phenomena.

Understanding of the project as the complex dynamic system among others requires operational and structural thinking. The presented model allows simulating the stressed phenomena in previous research, enriched by a clear interpretation of causes and effects and loop dominance. The chosen scenarios are illustrational and selected to present the common situations in project management. Nevertheless, the initial settings could be altered and the variety of simulation scenarios is very high.

The behaviour of the system grows from its structure. The demonstration of the structure and the simulation model was the goal of the paper. However, for the next step in our research, we plan to parametrise the model on the real data by Powell optimisation [6], [20] and categorise the possible scenarios of project development according to the findings from parametrisation. The parametrisation on the real projects' data also allow the sensitivity analysis and identification of the leverage points for the project management. Another possible extension consists in the implementation of the structure that will allow the simulation of the non-constant level of the resources throughout the project realisation.

References

[1] Barry, E. J., Mukhopadhyay, T., Slaughter, S. A. (2002). Software Project Duration and Effort: An Empirical Study. *Information Technology and Management*, *3*, 113–136.

[2] Bartoška, J., Kučera, P., Šubrt, T. (2015). Modification of the BCWS by Work Effort. In Martinčík, D., Ircingová, J., Janeček, P. (Eds.), *33rd International Conference Mathematical Methods in Economics* (pp. 19–24). Cheb: Czech Society for Operations Research.

[3] Bartoška, J., Kučera, P., Šubrt, T. (2016). Modification of the EVM by Work Effort and Student Syndrome phenomenon. In Kocourek, A., Vavroušek, M. (Eds.), *34th International Conference Mathematical Methods in Economics* (pp. 25–30). Liberec: Czech Society for Operations Research.

[4] Bartoška, J., Šubrt, T. (2012). The effect of human agent in project management. *Central European Journal of Operations Research*, *20*, 369–382.

[5] Coyle, R. G. (1996). *System Dynamics Modelling: A Practical Approach*. Boca Raton, FL: Chapman & Hall.
[6] Dangerfield, B., Roberts, C. (1999). Optimisation as a statistical estimation tool: An example in estimating the AIDS treatment-free incubation period distribution. *System Dynamics Review*, *15*, 273–291.
[7] Ford, D. N., Lyneis, J. M., Taylor, T. R. B. (2007). Project controls to minimize cost and schedule overruns: a model, research agenda, and initial results. In Sterman, J., Rogelio, O., Langer, R. S., Rowe, J.I., Yanni J.M., (Eds.), *25th International System Dynamics Conference*. Boston, MA: System Dynamics Society.
[8] Forrester, J. W. (1994). System Dynamics, Systems Thinking, and Soft OR. *System Dynamics Review*, *10*, 245–256.
[9] Goldratt, E. M. (1997). *The Critical chain*. Great Barrington, MA: North River Press.
[10] Jalili, Y., Ford, D. N. (2016). Quantifying the impacts of rework, schedule pressure, and ripple effect loops on project schedule performance. *System Dynamics Review*, *32*, 82–96.
[11] König, C. J., Kleinmann, M. (2005). Deadline Rush: A Time Management Phenomenon and Its Mathematical Description Relationships Between Critical Thinking and Attitudes Toward Women's Roles in Society. *The Journal of Psychology*, *139*, 33–45.
[12] Kučera, P., Bartoška, J., Šubrt, T. (2014). Mathematical Models of the Student Syndrome for Use in Project Management. In Marian, R., Gezik, P., Martin, L. (Eds.), *Proceedings of the International Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVII* (pp. 128–133), Virt: Slovak Society for Operations Research.
[13] Lyneis, J. M., Cooper, K. G., Els, S. A. (2001). Strategic management of complex projects: A case study using system dynamics. *System Dynamics Review*, *17*, 237–260.
[14] Meadows, D. H. (2008). *Thinking in systems: a primer*. Wright, D. (Ed.) White River Junction, Vt.: Chelsea Green Pub.
[15] Mildeová, S. (2013). Research Problem Description and Definition: From Mental Map to Connection Circle. *Journal on Efficiency and Responsibility in Education and Science*, *6*, 328–335.
[16] Özdamar, L., Alanya, E. (2001). Uncertainty Modelling in Software Development Projects (With Case Study). *Annals of Operations Research*, *102*, 157–178.
[17] Park, M., Pena-Mora, F. (2003). Dynamic change management for construction: Introducing the change cycle into model-based project management. *System Dynamics Review*, *19*, 213–242.
[18] Parkinson, C. N. *Parkinson's Law and Other Selected Writings On Management*. Singapure: Federal Publications.
[19] Pierson, K., Sterman, J. D. (2013). Cyclical dynamics of airline industry earnings. *System Dynamics Review*, *29*, 129–156.
[20] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. New York, NY: Cambridge University Press.
[21] Richmond, B. M. (1993). Systems Thinking: Critical Thinking Skills for the 1990s and beyond. *System Dynamics Review*, *9*, 113–133.
[22] Rodrigues, A., Bowers, J. (1996). System dynamics in project management: A comparative analysis with traditional methods. *System Dynamics Review*, *12*, 121–139.
[23] Sterman, J. D. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.
[24] Šviráková, E. (2014). System Dynamics Methodology: Application in Project Management Education. In Houška, M., Krejčí, I., Flégl, M. (Eds.), *11th International Conference on Efficiency and Responsibility in Education* (pp. 813–822). Prague: Czech University of Life Sciences Prague.
[25] System Dynamics Society. (2018). *What is System Dynamics* [online], Available: https://www.systemdynamics.org/what-is-sd [20 Mar2019].

Czech Tourism Demand: Modelling and Testing in VARs

Sára Bisová[1]

Abstract. In available analyses of tourism in the Czech Republic multivariate macroeconometric models are poorly applied. Usually only statistics summaries and graphical representations are published. This study offers Czech tourism analysis which focuses on investigation of relations and influences between tourism demand and the set of selected macroeconomic indicators. This set includes aggregates representing the economic situation of the analyzed country and also foreign and world indicators that are, under our assumptions, significantly connected with the decision making of the potential tourists (tourism demand) traveling to Czech destinations due to change in costs of their trips. These factors are: oil price, exchange rate and price level in the Czech Republic. For the purposes of mentioned analysis bivariate and multivariate vector autoregression models are specified, Granger causality tests are evaluated and impulse response functions are constructed. The results correspond with our assumption and suggest that an intervention or a shock influencing stated factors can significantly affect the tourism demand.

Keywords: Granger causality, impulse response functions, macroeconomic analysis, tourism demand, vector autoregression model.

JEL Classification: C32, C51, E52, E58
AMS Classification: 62M10, 91B84

1 Introduction

Tourism is one of the main sources of income and employment for many countries. This economic sector is very dynamic and is experiencing a continuing growth during the last decades and therefore collecting and analyzing tourism data becomes more and more relevant and desirable. According to UNWTO[2], business volume of tourism is comparable to that of oil exports, food products or automobiles and significantly influences many other economic sectors. In the Czech Republic, most of the analyses and studies are usually based on evaluating the progress of specific indicators or on benchmarking of selected territorial units by monitoring descriptive statistics and time series progression while more complex and sophisticated econometric analyzes are very uncommonly used in this area. We can define *tourism* as the activity or set of activities of people (visitors) taking a trip to a destination outside their usual environment, for less than a year, for an main purpose, including leisure, business and other purposes, other than to be employed by a resident entity in the place visited [6]. The basic tourism typology includes three forms of tourism [6] – domestic, inbound and outbound tourism. *Domestic tourism* includes activities of resident visitors within the country of reference either as a part of a domestic trip or part of an outbound trip. *Inbound tourism* includes activities of non-resident visitors within the country of reference on an inbound trip. *Outbound tourism* is then represented as activities of a resident visitor outside the country of reference, either as a part of an outbound or a domestic trip. By joining these three forms in different ways we can derive other forms of tourism [6]: *internal* (domestic and inbound tourism), *national* (domestic and outbound tourism) and *international* (inbound and outbound tourism). In this study we focus primarily on inbound tourism to the Czech Republic. In some cases we calculate also with domestic tourism that is not prime type of our interest, as discussed below.

There are many factors that determine tourism demand and a lot of them are widely discussed in tourism literature. Among commonly mentioned determinants belong changes in price level, oil prices, exchange rates, economic growth, exports etc. If we want to analyze the tourism demand we have to decide which indicator to choose for this purpose. When we look into recent studies indicators as: tourist arrivals (for example Huang, Silva, Hassani [4]), Song, Li [8], Wong, Song, Witt, Wu [9]) tourist expenditures (Song, Li [8]) or tourism receipts/revenues (Caglayan, Sak, Karymshakov [2]), nights spent (Biagi, Pulina [1]), etc. are applied. Huang, Silva, Hassani in [4] analyzed the causal relationship between oil price and tourist arrivals using data from US and nine European countries to explain the impacts of oil price volatility on tourist arrivals across countries.

[1] University of Economics Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, xbiss00@vse.cz.
[2] UNWTO - The United Nations World Tourism Organisation.

They conducted data-driven research using advanced nonparametric causality technique - Convergent Cross Mapping method (CCM) - and compared the results from this approach with two empirical causality methods which come under the time domain and frequency domain criteria. Applying the CCM method they proved the existence of one-directional causality running from oil prices to tourism arrivals for all countries while the comparing methods did not detect the same. Song, Li [8] specified model for tourism demand and construct ex post forecast for evaluation the quality of the model using two definitions of tourism demand - real tourist expenditures and arrivals in their aggregate and per capita form. They used Hong Kong data and three key source markets - visitors from Australia, the UK and the USA. To model the tourism demand they employed macroeconomic indicators including real GDP, CPI and exchange rate and defined ADL (autoregressive distributed lag) models for each of the three incoming countries. Their findings suggest that using tourist arrivals as demand for tourism is more influenced by origin country income and habit effects, while tourist expenditures are connected with destination prices relative to those in the origin country. Wong, Song, Witt, Wu [9] used more approaches for modelling and forecasting tourism demand for Hong Kong with effort to examine the efficiency of combining more forecasts using ARIMA, ADLM, ECM and VAR models. They employed tourist arrivals from key source markets (ten countries) as the tourism demand for Hong Kong. They conclude that although the combined forecasts do not always outperform the single model forecasts, they can reduce risk of forecasting failure and should be preferred to single model forecasts in many practical situations. Biagi, Pulina in [1] used VAR models and Granger causality in empirical analysis of Sardinia. Authors defined several models and conclude that the tourism supply is demand-driven, the tourism demand is quality-driven and there is simultaneous Granger causality between tourism demand and capacity (number of beds). Caglayan, Sak, Karymshakov [2] implemented three-stage panel Granger causality analysis to investigate the relationship between tourism revenue and gross domestic product using panel data of 135 countries divided into 11 geographical groups. The results differ between the groups – the explanation of non-existence of causality between analyzed variables was found for groups, where tourism sector is underdeveloped and brings small share in economy of countries.

For our study we use three forms of tourism demand: arrivals of non-residents and nights spent by non-residents at tourist accommodation establishments and occupancy rate of bed-places and bedrooms in hotels and similar accommodation in Czech Republic. Our aim is to check the relevancy of the three versions of tourism demand and to reveal the relationships between tourism demand and a set of factors which we assume influence the tourism demand. These factors are namely: oil price[3], exchange rate and price level of target country under the assumptions that the changes in factors influence tourists decisions about travelling to the country due to change in costs of their trips (increase in oil prices and Czech price level rise traveling costs and lowers the tourism demand while increase in exchange rate in national currency (CZK/EUR), depreciation of CZK, lowers the traveling costs and causes rise in tourism demand). We estimate set of bivariate VAR models for Granger causality (GC) testing and consequently we specify set of more-dimensional VAR models with aim to construct impulse response functions which can be also used for a purpose of causality testing and inform us, in comparison with GC testing, about the signs of the relations and influences between analyzed variables. The paper is organized as follows - section 2 states theoretical background of applied econometric approach, data are described in section 3, empirical analysis is presented in section 4 and conclusions are summarized in section 5.

2 VAR models

The structural form of VAR model (SVAR) without intercepts can be defined as follows [5]

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{B}\mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t$ is a structural disturbance vector generated by vector white noise process, with identity covariance matrix $E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}) = \mathbf{\Lambda}$, where diagonal elements are variances of structural shocks, $\mathbf{\Pi}(L)$ is a polynomial matrix in the lag operator, $\mathbf{y}_t$ is a vector of *m* endogenous variables and **A** is a matrix of (*m* x *m*) parameters representing simultaneous relations between endogenous variables. The elements outside the diagonal in matrix **B** may be non-zero, therefore, some of the shocks can influence more endogenous variables of the system.

We can estimate the reduced form [5]

$$\mathbf{y}_t = \mathbf{A}^{-1}\mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{v}_t, \tag{2}$$

[3] Crude oil is one of the world´s most important energy source and therefore influence all the national and international economic sectors.

where $\mathbf{v}_t$ is a disturbance vector of reduced form with a covariance matrix $E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{\Sigma}$. For the disturbance vectors $\mathbf{u}_t$ and $\mathbf{v}_t$ holds the following formulae [5]

$$\mathbf{v}_t = \mathbf{A}^{-1}\mathbf{B}\mathbf{u}_t \quad \text{or} \quad \mathbf{A}\mathbf{v}_t = \mathbf{B}\mathbf{u}_t \tag{3}$$

and

$$E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1} \quad \text{or} \quad \mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}. \tag{4}$$

The matrix $\mathbf{A}$ identifies the relation between the structural disturbances $\mathbf{u}_t$ and the reduced form shocks $\mathbf{v}_t$. We employ the *recursive identification* scheme, so-called *Choleski decomposition*, to get a just identified VAR, where the matrix $\mathbf{A}$ is lower triangular and the matrix $\mathbf{B}$ is diagonal, see [5]

2.1 Causality testing in VARs

We employ two approaches to investigate the causality: Granger causality tests and impulse response functions, briefly described in sections 2.2 – 2.3 below.

2.2 Granger causality

Granger concept of causality (GC), presented in Granger [3], is commonly applied when working with VAR models. The results of this test inform us, if the past values of one variable (or the subset of variables) can help in predicting the remaining variable. In two-variable (*X*, *Y*) system we can test, if *X* can be better predicted using past values of *Y* compared to using only historical data of *X*. In this most simple situation it is a partial *F*-test on regression parameters of lagged *Y* data in the regression for *X*. More about GC test in Granger [3] or Hušek [5].

2.3 Impulse response functions

The estimation of VAR models is usually followed by construction of impulse response functions (IRF). IRF can inform us about the signs of relationships and influences between analyzed variables and therefore can be also used for causality analysis. IRFs inform us about the response of all variables of the VAR system to a unit exogenous shock in each endogenous variable. For deriving and interpretation of IRF see Sims [7].

3 Data

Dataset collected for our analysis consists of six time series:

- three forms of tourism demand[4]:
  - *OCCUPANCY* - net occupancy rate of bed-places and bedrooms in hotels and similar accommodation in Czech Republic;
  - *NIGHTS* – total nights spent by non-residents at tourist accommodation establishments[5];
  - *ARRIVALS* - arrivals of non-residents at tourist accommodation;
- *CPI* - Czech consumer price index with basis in 2015 (2015=100) [6];
- *ER* – Euro/ECU exchange rate in national currency[7];
- *OIL_PRICE* – Brent crude oil prices (Europe) in dollars per barrel[8].

We use *NIGHTS* and *ARRIVALS* of *non-residents* while we assume that inbound tourism is most affected by the change in oil price and exchange rate. The main part of inbound tourism to Czech Republic consists of EU countries[9] and therefore Euro/ECU exchange rate was chosen as a representative exchange rate for the purpose of our analysis. Brent crude oil is the most important type of crude oil in Europe, so we use its data for oil prices. All the time series are measured at quarterly frequencies from Q1 2002 to Q4 2018 (68 observations) because earlier data of Czech indicators for tourism demand indicators are not available. Figure 1 shows times series plots of all the variables listed above in levels - the non-stationarity is evident.

[4] The source of all the time series for tourism demand is Eurostat - http://ec.europa.eu/eurostat.
[5] It includes - hotels, holiday and other short-stay accommodation, camping grounds, recreational vehicle parks and trailer parks.
[6] Czech statistical office – http://www.czso.cz .
[7] Eurostat - http://ec.europa.eu/eurostat.
[8] FRED (Federal reserve bank of St. Louis) – http://fred.stlouisfed.org.
[9] For example around 20% of Czech inbound tourism represents Germany.

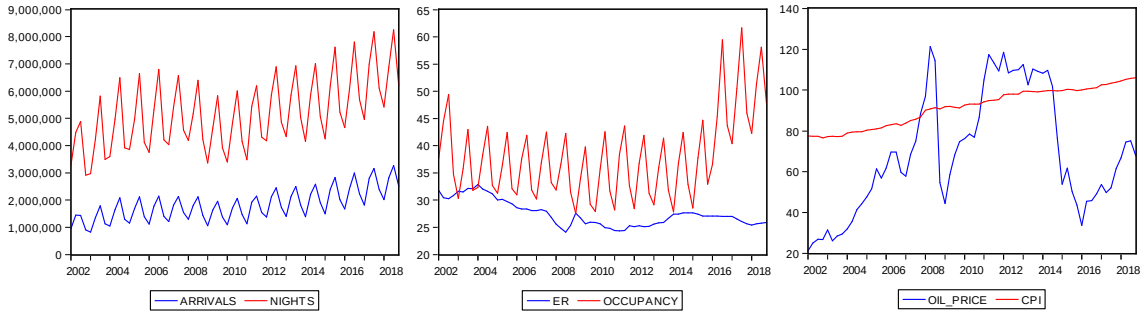


Figure 1 Time series plot

The tourism demand time series include seasonality. For Granger causality testing the data were not seasonally adjusted, because seasonality brings important information to this test. For impulse response functions (IRFs) analysis of more-dimensional models tourism demand variables and *OIL_PRICE* were seasonally adjusted to reach more correct specification of the models and better interpretable IRFs that are quicker returning to equilibrium. For the estimation EViews 6 and GRETL 1.9.5 software were used.

Table 1 shows the ADF tests of all the time series mentioned above in their level form. All the variables are non-stationary at 5 % significance level and therefore should be transformed to reach the stationarity for the application in VAR models. In order to eliminate the non-stationarity we use differences of logarithms (they can be approximatively interpreted as growth rates). All the time series are stationary in first differences of their logarithms (abbreviation d_log in fourth column) as shown in table 1. *OCCUPANCY* was only differentiated, because it is measured in percentage terms and therefore may not be logarithmic before the differentiation.

level	*t*-statistic	Prob.*	d_log	*t*-statistic	Prob.*
ARRIVALS	0.936802	0.9954	***ARRIVALS***	-4.087485	0.0021
CPI	-0.379915	0.9061	***CPI***	-3.378008	0.0152
ER	-2.207075	0.2057	***ER***	-5.922445	0.0000
NIGHTS	-0.111735	0.9430	***NIGHTS***	-3.792746	0.0049
OCCUPANCY	-0.656105	0.8497	***OCCUPANCY***	-3.683449	0.0067
OIL_PRICE	-2.040144	0.2694	***OIL_PRICE***	-5.942456	0.0000

*MacKinnon (1996) one-sided p-values
Notes: Test critical values -3.54 (1% level), -2.91 (5% level), -2.59 (10% level)

Table 1 ADF tests for variables in levels and d_log transformation

4 Application and results

4.1 Granger causality testing

We focused on the Granger causality test between: Czech price level and tourism demand variables, exchange rate and tourism demand and oil price and the set of tourism demand variables. The results are available in Table 2 - the first and fourth columns contain tested pair of variables in given order, second and fifth columns inform us about the number of included lags in the appropriate bivariate VAR and the third and sixth columns include the p-value for the test.

Almost all the results correspond with our assumptions. All the pairs of variables have uni- or bi-directional GC at 5 % significance level, except relation between *ER* and *OCCUPANCY*, where was not found any positive GC test in any direction. We have to say, that occupancy rate includes two types of tourism – inbound tourism and also domestic tourism. Domestic tourism obviously doesn´t react to changes in exchange rate so strongly. Moreover, inbound tourism includes also tourists from other countries than the EU. These facts weaken the relation between variables *ER* and *OCCUPANCY* and presumably weaken the results of GC tests, and impulse response functions as well. According to results of GC tests there is simultaneous relationship at 5 % significance level between price level and the tourism demand in all definitions, only GC test running from *CPI* to *OCCU-*

PANCY at only 10% significance level. Within these simultaneous relationships we can interpret the influence running from tourism demand variables to price level so that increasing prices in hospitality management projected into *CPI* are a result of the higher demand for accommodation capacity and for the related services. For *ARRIVALS* and *NIGHTS* there is positive GC test running from *ER* to these two definitions of tourism demand. According to our assumptions there are positive GC tests running from *OIL_PRICE* to tourism demand in all three forms at 5% significance level. Within the interpretations we must naturally keep in mind that the outputs of GC tests cannot be interpreted as causality in economic sense of cause and effect, this causality is based on principle of predictions.

GC test	df*	prob.	GC-test	df*	prob.
CPI › ARRIVALS	6	0.0000	***ARRIVALS › CPI***	6	0.0029
CPI › NIGHTS	6	0.0000	***NIGHTS › CPI***	6	0.0003
CPI › OCCUPANCY	5	0.0605	***OCCUPANCY › CPI***	5	0.0043
ER › ARRIVALS	7	0.0058	***ARRIVALS › ER***	7	0.1625
ER› NIGHTS	7	0.0339	***NIGHTS › ER***	7	0.1173
ER › OCCUPANCY	4	0.6013	***OCCUPANCY › ER***	4	0.5971
OIL_PRICE › ARRIVALS	7	0.0173	***ARRIVALS › OIL_PRICE***	7	0.4068
OIL_PRICE› NIGHTS	7	0.0089	***NIGHTS › OIL_PRICE***	7	0.5030
OIL_PRICE › OCCUPANCY	7	0.0005	***OCCUPANCY › OIL_PRICE***	7	0.3247

*Lag order selected by the AIC.

Table 2 Selected Granger causality tests

4.2 Impulse response functions

We estimated three four-variable VAR(3) models, although in some of them more than 3 lags were recommended by AIC criterion, because IRFs for 4 and more lagged VAR models included worse interpretable outputs with responses changing signs. Figures 2 and 3 show IRFs representing responses of the tourism demand variables (in rows) to unit exogenous shocks in *ER*, *OIL_PRICE* and *CPI* (in columns).

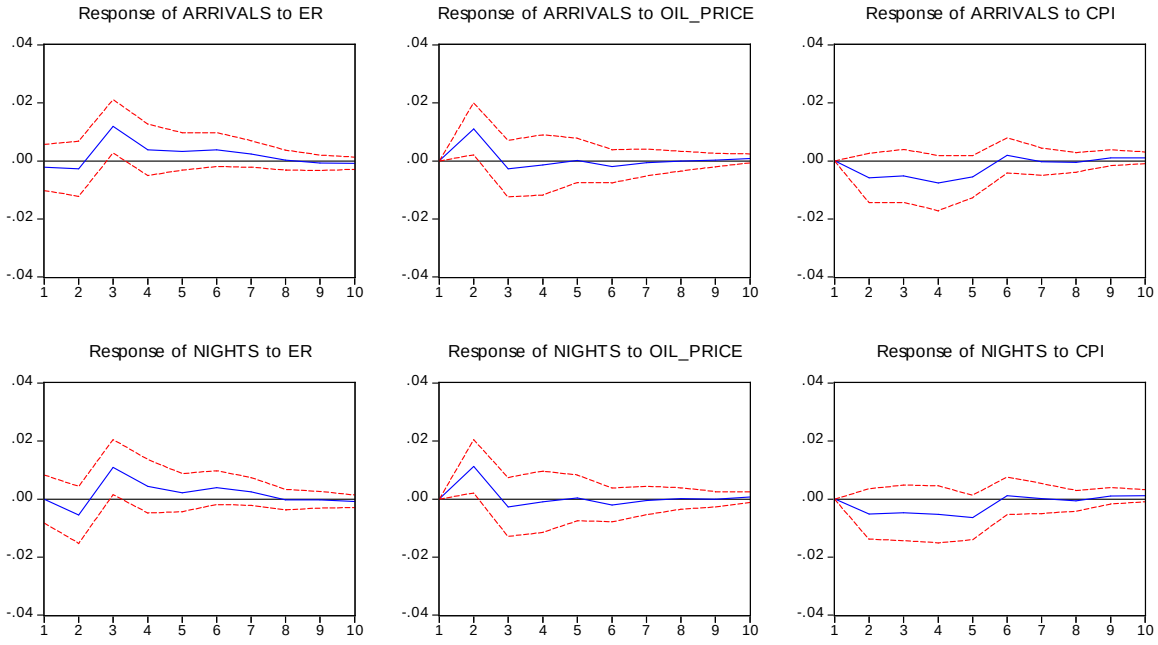


Figure 2 Impulse response functions – part 1 (response to Cholesky one S.D. innovations ± 2 S.E.)

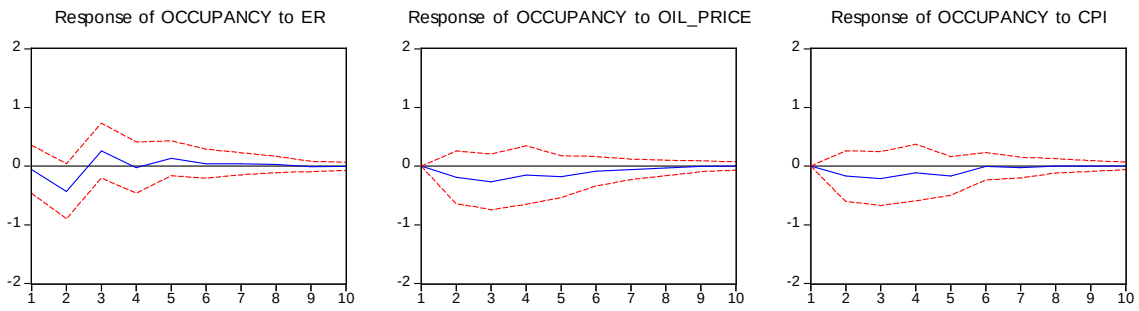


Figure 3 Impulse response functions part 2 (response to Cholesky one S.D. innovations ± 2 S.E.)

The IRFs are very similar for *ARRIVALS* and *NIGHTS* and slightly different for *OCCUPANCY* - it probably reflects the fact that this variable includes also domestic tourism of residents. The exogenous shock in exchange rate (depreciation) raises all the tourism demand variables with a slight persistence at the beginning of the simulated horizon, the weakest reaction has *OCCUPANCY* – it corresponds with the same problem in GC caused very likely by the domestic tourism included in this indicator, which presumably does not significantly react to changes in exchange rate. *OCCUPANCY* decreases immediately as a response to a positive shock in oil prices while the reactions of *ARRIVALS* and *NIGHTS* are delayed again. Unit positive shock in *CPI* is followed by decrease in Czech tourism demand in all the three forms.

5 Conclusion

In this paper three forms of tourism demand were applied to analyze the relations between tourism demand and three important determinants of tourism demand in general – oil prices, exchange rates and price level of analyzed country. The analysis was performed for Czech Republic. We employed Granger causality tests in bivariate and impulse response functions analysis in more-dimensional VAR models. We conclude that all the definitions of tourism demand are relevant, but occupancy rate is the poorest one for analyzing the relation of tourism demand and exchange rate CZK/EUR. The IRFs outputs match our assumptions (excepting initial persistence in some cases for the first simulated periods) about the relations between the variables. Our analysis suggests that an intervention or a shock influencing oil prices, exchange rates and price level can significantly affect the tourism demand. Breitung-Candelon test for detecting short- and long-run predictability of the system and SVAR models are the subjects of our upcoming analyses.

Acknowledgements

The research project was supported by Grant No. IGA F4/78/2018, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Biagi, B., Pulina, M. (2009). Bivariate VAR models to test Granger causality between tourist demand and supply: Implications for regional sustainable growth. *Paper in Regional Science*, 88(1), 231-244.
[2] Caglayan, E., Sak, N., Karymshakov, K. (2012). Relationship between tourism and economic growth: a panel Granger causality approach. *Asian Economic and Financial Review*, 2(5), 591-602.
[3] Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424-438.
[4] Huang, X., Silva, E., Hassani, H. (2018). Causality between Oil Prices and Tourist Arrivals. *Stats*, 1(1), 134-154.
[5] Hušek, R. (2007). *Ekonometrická analýza.* Praha: Oeconomica.
[6] Methodological manual for tourism statistics - Version 3.1 – 2014 edition. *Eurostat Manuals and Guidelines*, European Union. Available online [2.3.2019]: http://ec.europa.eu/eurostat/.
[7] Sims, Ch. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1-48.
[8] Song, H., Li, G. (2010). Tourism demand modelling and forecasting: how should demand be measured? *Tourism Economics*, 16(1), 63-81.
[9] Wong, K., Song, H., Witt, S. F., Wu, D. C. (2007). Tourism forecasting: to combine or not to combine? *Tourism Management*, 28, 1068-1078.

The Persistence of Consumption and Disposable Income in the V4 Countries

Terézia Bobošová[1], Luboš Střelec[2]

Abstract. This paper deals with the measurement and comparison of the persistence of aggregate private consumption and aggregate disposable income in the Visegrad Group countries (Czech Republic, Hungary, Poland and Slovakia). The analysis is done by using a non-parametric method applied to quarterly data covering the period from 1995(1) to 2017(4). It is concluded that the aggregate consumption in Slovakia is more persistent than in Poland. When the degree of persistence is high, a potential shock tends to have a more long-lasting effect. The consumption expenditures in the Czech Republic are more persistent than the disposable income in contrast to Poland, where disposable income is much smoother than consumption.

Keywords: consumption, disposable income, persistence, time series, Visegrad Group (V4) countries.

JEL Classification: C01, C22, E10, E12
AMS Classification: 62M10, 62P20, 91B42, 91B55, 91B84

1 Introduction

This paper addresses the issue of the degree of persistence of aggregate consumption expenditures and aggregate disposable income in V4 countries and engages its economics policy implication. On the one hand, persistence can reduce the negative incidence and length of unexpected economic conditions. On the other hand, persistence is particularly problematic for the effectiveness of the stated countercyclical policies acting through consumption.

The paper is organized as follows: section 2 brings a brief overview of the theoretical approach to the persistence of aggregate consumption expenditures and aggregate disposable income and the recent literature discussion in the context of the macroeconomic policies. Section 3 introduces data and methods. The results of the empirical analysis are presented in section 4. Section 5 is dedicated to conclusions of the analysis.

2 Theoretical Background

Milton Friedman's Permanent Income Hypothesis (1957) and Franco Modigliani and Richard Brumberg's (1954) analogous Life-Cycle Model were 'new' consumption function theories of the 1950s, which have brought an influential contribution to answering the question how consumption, and thus savings, depends on the development of income.

A core statement of these hypotheses originates from the basic intuition that individuals would like to smooth consumption over time and not let it fluctuate with short-run fluctuations in income. Consumption is determined by the value of lifetime resources; it is determined by permanent income. Individuals consume a fraction of their permanent income in each period and thus the average propensity to consume would equal the marginal propensity to consume. The propensity itself could vary with several factors.

So how much of their income people consume in the various time periods? This is an important issue in macroeconomics – the flip side of total consumption is total savings, and the evolution of savings over time in a country is important to its capital formation and business cycles.

First, since permanent income depends, not just on income in one year, but on income over a number of years, perhaps all of life, it is unlikely to fluctuate much in response to short-term fluctuations in income. Permanent income is expected to be smoother than measured or current income. Current income appears to be more

[1]Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: terezia.bobosova@mendelu.cz.
[2]Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: lubos.strelec@mendelu.cz.

volatile than consumption, implying that individual consumers don't respond to fluctuations that are not perceived as permanent.

Belbute and Caleiro (2009) consider aggregate consumption to be the most stable component of aggregate demand that means the consumption is less volatile than GDP. The above-mentioned theories (The Life-Cycle Theory, Permanent Income Theory and Random Walk Hypothesis) explain this pattern by assuming that households try to smooth their lifetime consumption. In particular, the Hall's (1978) Random Walk Theory suggests that future income is correctly anticipated and incorporated into permanent income so that current consumption fully reflects this information. Since all these expectations in future circumstances are already taken into account, only unexpected shocks can alter wealth and thus consumption. Put differently, in accordance with those theories, consumption displays some degree of inertia or persistence (Belbute and Caleiro, 2009).

Willis (2003) explicates the persistence as a measure of the speed with which a variable returns to its baseline after a shock, i.e. some event (for instance, a macroeconomic policy measure) that provoked an increase (or decrease) in that variable. In this sense, when the degree of persistence is small, a shock tends to have more temporary effects and conversely when the degree of persistence is high, a shock tends to have more long-lasting effects (Belbute and Caleiro, 2013). Following a shock, using a less persistent variable can stabilize economic issue in a shorter time. (Dias and Marques 2010). This is particularly problematic for the effectiveness of the present countercyclical policies that act through consumption.

The first macroeconomic studies incorporating the issue of persistence has at first focused on inflation and not on real variables such as consumption. Studies aiming at the evaluating persistence of inflation are for example Marques (2004), Huang and Liv (2002), Ascari (2003), Willis (2003), Levin and Piger (2004), Gadzinski and Orlandi (2004), Cogley and Sargent (2001, 2008), Pivetta and Reis (2007) and Fuhrer (2009). These studies evaluate the role of staggered wage-setting and staggered price-setting as a source of persistent real effects of monetary shocks.

Multiple studies investigate the persistence of others macroeconomic variables, such as aggregate output or unemployment, often also with connection with monetary policy shocks. The persistence of shocks to aggregate output has been the subject to investigation in various empirical analyses, for example Kiley (2000), Maury and Tripier (2003), Bouakez and Kano (2006) and Wang and Wen (2006).

The macroeconomic policies play the dominant role in smoothing the business cycle. For assessing the effectiveness of economic policy measures it is important to quantify the response of consumption to a shock. Quantifying the response of consumption to a shock implies evaluating the persistence of consumption. It may allow setting the right timing of those measures in order to overwhelm the harmful effect of a shock over consumption (Belbute and Caleiro, 2009).

The Permanent Income Hypothesis (Friedman, 2015) suggests that households consume their permanent income. Temporary shocks need to be smoothed by saving/borrowing because permanent income differs from current income. But for highly persistent shocks, current income is similar to permanent income, so simply spending current income is close to optimal (Mendes and Pennings, 2017).

The literature applied to the persistence of consumer's habits has recently also gained some relevance in psychology and marketing. Belbute and Caleiro (2009) try to explain how the behaviour of consumers in a country with specific psychosocial consumption habits may lead to the persistence of consumption at an aggregate level.

3 Data and Methods

The research is based on the time series analysis using Visegrad Group countries data covering the period from the first quarter of 1995 till the last quarter of 2017. Eurostat is the main data source for all used time series. The time series of disposable income is available only at an annual frequency. The quarterly path is estimated on the basis of external quarterly information of a relevant economically related variable. The real gross domestic product is used in this paper as a quarterly indicator series for disaggregation of annual figures to quarterly figures. Because of the time consistency, quarterly values have to match annual values (the sum of quarterly values of the GDP must be equal to the annual value). The correlation coefficient between annual time series of real gross domestic product and annual time series of disposable income reaches value 0.99 in the case of the Czech Republic; 0.98 in Hungary; 0.99 in Poland and 0.99 in Slovakia. The quarterly time series of GDP are aggregated into annual data and then a proportion of each quarter value to annual value is computed. Following these proportional shares, annual values of disposable income can be divided into quarterly values.

The time series of disposable income is measured as per capita aggregate in euro currency and seasonally adjusted by the TRAMO/SEATS procedure. For acquiring the real terms of disposable income, the GDP implicit deflator is used. The GDP implicit deflator is calculated by dividing an aggregated GDP measured in current prices by the same aggregate measured in prices of the year 2010.

The second essential time series is the time series of consumption expenditures. The quarterly time series of consumption is also expressed in real terms due to the same GDP deflator usage and measured as per capita aggregate in euro currency and seasonally adjusted by the TRAMO/SEATS procedure.

Marques (2004) and Dias and Marques (2010) have suggested a new non-parametric measure of persistence denoted by γ, based on the relationship between persistence and mean reversion. The definition of this procedure is the unconditional probability of a given series not crossing its mean in period t. The statistic γ is then characterized by the following expression:

$$\gamma = 1 - \frac{n}{T},$$

where n stands for the number of times the series crosses the mean during a selected period T and the ratio $\frac{n}{T}$ provides the degree of mean reversion. The usage of this non-parametric procedure is appropriate because of non-stationary time series of consumption expenditures and disposable income. The high persistence indicates that after a shock the series reverts to its mean more seldom (higher value of γ), by contrast low persistence denotes that after a shock the series crosses its mean more frequently (lower value of γ).

4 Results

The non-parametric method, which measures the degree of mean reversion, was used to evaluate the persistence of consumption and disposable income in V4 countries. In order to compute the mean of each series the Hodrick-Prescott (HP) filter was used following Marques' suggestion (2004) that a time varying mean is more appropriate than the simple average. For the calculation of the degree of mean reversion the cyclical component extracted from used Hodrick-Prescott filter is applied. Figures 1–4 show the cyclical components of consumption and disposable income for the V4 countries together with the outlined zero average.

Table 1 summarizes the results of persistence of consumption expenditures and disposable income in the V4 countries.

Country	Consumption expenditures		Disposable income	
	n	γ	**n**	γ
Czech Republic	13	0.8587	21	0.7717
Hungary	12	0.8696	12	0.8696
Poland	21	0.7717	11	0.8804
Slovakia	11	0.8804	9	0.9022

Table 1 The persistence of consumption expenditures and disposable income in the V4 countries

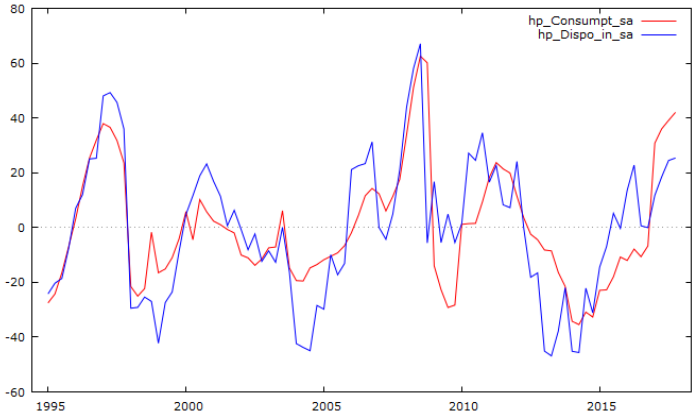


Figure 1 The cyclical component of aggregate consumption in Czech Republic

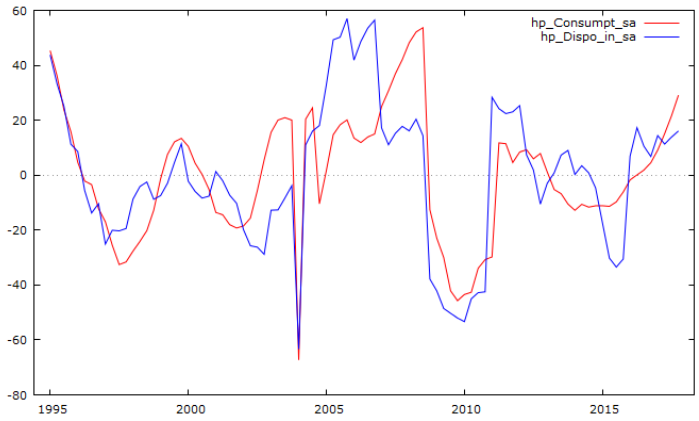


Figure 2 The cyclical component of aggregate consumption in Hungary

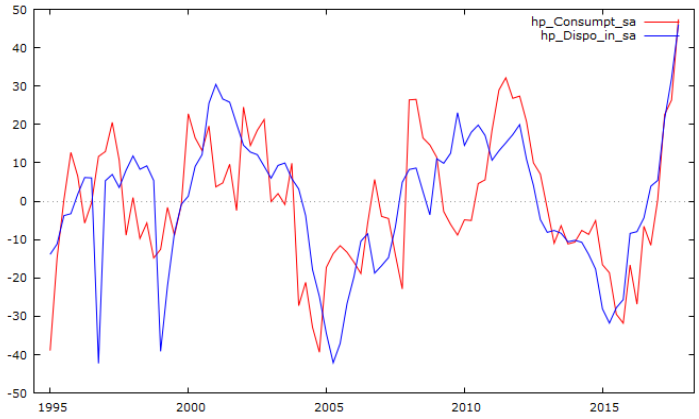


Figure 3 The cyclical component of aggregate consumption in Poland

Figure 4 The cyclical component of aggregate consumption in Slovakia

The non-parametric measure denoted by γ shows similar conclusions for the persistence of consumption expenditures in the Czech Republic, Hungary and Slovakia. Consumption expenditures in Poland are less persistent, which means that after an unexpected change (a shock) consumption deviates quickly from its trend and tends to stay above or below it less time than in the Czech Republic, Hungary or Slovakia. In the context of economic forecasting, especially over a short run, the knowledge of persistence helps to indicate whether consumption will stay away from its trend after a shock.

In contrast, policy innovation or a random shock that affects household expenditures will tend to have more permanent effects on consumption in the Czech Republic, Hungary and Slovakia. Countercyclical policies that operate through consumption can be ineffective if consumption is very persistent. To put it in another way, the same fiscal stimulus would be more effective in Poland than in the other three countries. Polish consumption will stay above/below its trend less time than in other V4 countries.

The surprising result is that the disposable income in Poland shows considerably lower volatility than consumption. Consumption would be expected to be smoother than disposable income, just as the findings for the Czech Republic confirm. The Deaton's paradox admits a higher smoothness of the current income to the permanent income. The persistence of consumption expenditures and disposable income in Hungary and Slovakia are very similar.

5 Conclusions

Conclusions of this paper suggest that aggregate consumption in Poland is less persistent than in the Czech Republic, Hungary and Slovakia, so consumption expenditures in Poland will stay less time away from the trend after a shock. Furthermore, the relatively lower degree of persistence shown in the consumption in Poland can be beneficial for the effectiveness of the countercyclical fiscal and monetary policies. The persistence of consumption expenditures in the Czech Republic, Hungary and Slovakia is even higher than in the United Kingdom, Italy and Norway defined in (Belbute and Caleiro, 2013). The evidence for high persistence in aggregate consumption mainly in these three countries reflects strong habit formation mechanisms. However, given the high degree of persistence, consumption will react slowly there but policies that act trough consumption will tend to have long-lasting effects.

Acknowledgements

Supported by the grant No. PEF_DP_2019034 of the Grant Agency IGA PEF MENDELU.

References

[1] Ascari, G. (2003). Price/Wage Staggering and Persistence: A Unifying Framework. *Journal of Economic Surveys*, 17, 511–540.
[2] Belbute, J. & Caleiro A. (2009). Measuring the Persistence on Consumption in Portugal. *Journal of Applied Economic Sciences*, 2(8), 185-202.
[3] Belbute, J. & Caleiro, A. (2013). Cross Country Evidence on Consumption Persistence. *International Journal of Latest Trends in Finance and Economic Sciences*, 3, 440-455.
[4] Bouakez, H. & Kano T. (2006). Learning-By-Doing or Habit Formation? *Review of Economic Dynamics*, 9, 508–524.
[5] Cogley, T. & Sargent, T.J. (2001). Evolving post-WWII U.S. Inflation Dynamics. *NBER Macroeconomics Annual*, 16, 331–373.
[6] Cogley, T., Primiceri G. E. & Sargent T. J. (2008). Inflation-gap persistence in the U.S. *National Bereau of Economic Research*, Working Paper No. 13749.
[7] Dias, D. & Marques C. R. (2010). Using Mean Reversion as a Measure of Persistence, *Economic Modelling*, 27, 262-273.
[8] Friedman, M. (1957). *A theory of Consumption Function*. 1. ed. Princeton University Press. ISBN 0-691-04182-2.
[9] Friedman, M. (2015). *A theory of Consumption Function*. 1. ed. Martino Fine Books. ISBN 978-1614278122.
[10] Fuhrer, J. C. (2009). Inflation Persistence. Research Department Working Papers. *Federal Reserve Bank of Boston*, Research Department, Working Papers, 09–14.
[11] Gadzinski, G. & Orlandi, F. (2004). Inflation persistence in the European Union, the Euro Area and the United States. *European Central Bank*, Working Paper, 414.
[12] Kiley, M. (2000). Endogenous Price Stickiness and Business Cycle Persistence. *Journal of Money, Credit and Banking*, 32, 28–53.
[13] Levin, A. T. & Piger, J. M. (2004). Is inflation persistence intrinsic in industrial economies? *European Central Bank*, Working Paper, 34.
[14] Levin, A.T. & Piger, J.M. (2004). Is inflation persistence intrinsic in industrial economies? *European Central Bank*, Working Paper, 334.
[15] Marques, C. R. (2004). Inflation Persistence: Facts or Artefacts? *Working Paper*.
[16] Maury, T. P. & Tripier F. (2003). Output Persistence in Human Capital-Based Growth Models. *Economics Bulletin*, 5, 1–8.
[17] Mendes, G. & Pennings A. (2017). Consumption smoothing and shock persistence: optimal simple fiscal rules for commodity exporters. *World Bank Policy Research*, Working Paper, 8035.

[18] Modigliani, F. & Brumberg, R. (1954). Utility analysis and the consumption function: an interpretation of the cross-section data, in K. Kurihara (ed.). *Post-Keynesian Economics*. Rutgers University Press. New Brunswick, NJ, 388–436.
[19] Pivetta, F. & Reis, R. (2007). The persistence of inflation in the United States. *Journal of Economic Dynamics and Control*, 31, 1326–1358.
[20] Wang, P. & Wen, Y. (2006). Another look at sticky prices and output persistence. *Journal of Economic Dynamics and Control*, 30, 2533–2552.
[21] Willis, J. L. (2003). Implications of structural changes in the U.S. Economy for pricing behavior and inflation dynamics. *Economic Review*, Federal Reserve, Bank of Kansas City.

Estimation of Long-Term Projects in Terms of Uncertainty of Inputs

Simona Hašková[1]

Abstract. One of the criteria underlying the decision concerning acceptance or rejection of a project is the internal rate of return (IRR). The project is acceptable if the discount rate applied is lower than the IRR. We focus on the project of innovative commodity production whose consumption is of interest to the government. The need for a large initial investment and the high production costs do not stimulate a sufficient demand to guarantee an adequate IRR. Therefore, such a project is subsidized. Given the ambiguity about the amount of investment, future market price and demand, only the intervals of these values can be estimated. Concurrently, any relevant reason to prefer a particular value does not exist. We consider this situation as uncertain on the input side of the evaluation process, with the IRR on the output side being consequently uncertain. A tool to answer the question of whether the system of subsidies is adequate is the fuzzy approach. The electric car production project analysis shows that this approach provides superior information, which helps the investor e.g., to assess the degree of competitiveness of a project given to the alternative yield rate.

Keywords: Fuzzy logic, input uncertainty, internal profitability, interval analyses.

JEL Classification: C51, C58
AMS Classification: 90B50, 90B90

1 Introduction

One of the important criteria underlying the decision to accept or reject a long-term project is its internal rate of return (IRR). This is considered to be the annual discount rate, balancing the initial net capital expenditure with the present value of the total cash flow of the net annual income generated by the project, and guaranteeing that the investor achieves the required rate of return on capital [5]. IRR is seamlessly quantifiable when all annual incomes are non-negative. The project is acceptable to the investor if its IRR is not lower than the annual return rate of return on equally risky capital market investments [10].

Let us suppose the project that involves the production of an innovative commodity whose consumption is of interest to the government. However, the need for a large initial investment in the project and the high production costs do not stimulate a sufficiently high demand to guarantee adequate profitability. The government can stimulate the increase in IRR of the project by means of a system of various incentives to make it more attractive for investors. An example of such a commodity is the electric vehicle [13]. The national governments are implementing various incentive policies such as tax exemption, purchase subsidies, free parking and various driving privileges, which should have a positive impact on the growth in demand for electric cars.

Given the uncertainty about the future supported demand projecting into vagueness about the future market prices and production volumes, only the intervals of these input values of the project can be reliably estimated; furthermore, no relevant reason exists as to why a particular value within a given interval should be preferred [7]. Thus, we are in a situation of uncertainty on the input side of the evaluation process with the IRR on its output side as an uncertain parameter [8].

[1]The Institute of Technology and Business/The School of Expertness and Valuation, Okružní 517/10, České Budějovice, Czech Republic, e-mail: haskovas@post.cz.

The tool to answer the question of whether the system of subsidies offered by the government is beneficial from the point of view of investors is the algebra of intervals based on the herein presented fuzzy approach. This, unlike the conventional approach, also provides above-standard information about the potential danger of below-average rates of return that may help the investors in their decision-making.

For this purpose, the fuzzy linguistic evaluation will be discussed and applied as a relevant alternative to conventional management budgeting techniques. The methodological part will be devoted to the comparison of the conventional approach with the fuzzy approach. These approaches will be further demonstrated in the solution of the problem that the investors often face when deciding about investment in a subsidized commodity. Specifically, it deals with the IRR evaluation of the project of electric vehicle production, whose uncertain input, stemming from the prior ignorance of future demand, is the percentage of production margin. The advantages of the fuzzy approach are discussed and summarized in the final section along with the benefits of the analysis.

2 The methodological concept of the conventional and fuzzy approach to internal yield rate calculation

Let one of the conventional criteria for assessing the profitability of the investment in projects be the internal rate of return (IRR) criterion (see, e.g. [3]), which in our case is defined by the formula:

$$\sum_{i=1}^{n} CF_i \,/(1 + IRR)^i = -CF_0 \qquad (1)$$

in which input variables are:

- negative payment $CF_0 = -(I + NW)$, where I stands for *capital investment*, NW stands for *net working capital expenditure*,
- $CF_i = R_i - VC_i - FC_i - T_i$, $i > 0$, where R_i stands for *revenue from sales*, VC_i for *variable costs*, FC_i for *fixed costs* and T_i for *tax* in each year i of the lifetime of the given project.

The symbols I, NW, R_i, VC_i, FC_i and T_i, i = 1, 2,…, n, are parameters of input variables. Equation (1) is relevant and unambiguously solvable only in the cases where the numerators of all addends on the left side are nonnegative (for details see [2]). Let *IRR* be the wanted value of implicit output variable IRR of (1).

Suppose that some parameter x of any input variable is uncertain, i.e., only the interval $\langle x_{min}, x_{max}\rangle$ of all its possible values can be estimated with no other relevant information available. The question is: what value of this parameter should be inserted as a relevant input variable?

2.1 Conventional statistical approach

To answer the question, the conventional approach would rely on the *principle of indifference*, which says that if there are multiple alternative values for which we do not have any relevant reason to prioritize one over the other, then the same probability of occurrence for each will be assigned (in other words it says how an agent should respond in the absence of evidence [11]).

On the interval $U = \langle x_{min}, x_{max}\rangle \subset \Re$ the values can be regarded as values of the continuous random variable α given by constant probability density $f_\alpha(x) = 1 / (x_{max} - x_{min})$ on the interval U and the statistically expected value $E[\alpha] = \int_U (x / (x_{max} - x_{min})) \cdot dx = (x_{max} + x_{min}) / 2$; in the special case, when α is a constant, $E[\alpha] = \alpha$.

In this sense, the conventional approach replaces parameters I, NW, R_i, VC_i, FC_i, T_i with E[I], E[NW], $E[R_i]$, $E[VC_i]$, $E[FC_i]$, $E[T_i]$, i = 1,2,…,n, and also the equations $CF_0 = -(I + NW)$ and $CF_i = R_i - VC_i - FC_i - T_i$ with the equations $E[CF_0] = -(E[I] + E[NW])$ and $E[CF_i] = E[R_i] - E[VC_i] - E[FC_i] - E[T_i]$. The statistically expected values $E[CF_0]$, $E[CF_i]$ and E[IRR] are inserted to the relation (1) the result of which is the formula (2)

$$\sum_{i=1}^{n} \mathrm{E}[\mathrm{CF_i}] \,/(1 + \mathrm{E}[\mathrm{IRR}])^{\mathrm{i}} = \mathrm{E}[-\mathrm{CF_0}] \tag{2}$$

From formula (2) the value of E[IRR] can be calculated by an iterative method. An example of a particular procedure is given in [9], where the authors suggest and analyse two new two-step iterative methods for solving the system of nonlinear equations using quadrature formulas; the use of the GeoGebra program as a convenient tool for solving similar types of tasks by the method of numerical approximation is shown in [6].

2.2 Fuzzy approach

Contrary to the conventional approach, the fuzzy approach interprets the interval $U = \langle x_{\min}, x_{\max}\rangle \subset \mathbb{R}$ of all possible values of uncertain parameter x as the support of non-fuzzy subset $\underline{A} = \{(x, \mu_{\underline{A}}(x)): x \in \mathbb{R}\}$, $\mu_{\underline{A}}(x) = 1$ for $x \in U$, $\mu_{\underline{A}}(x) = 0$ otherwise, which is a fuzzy number [1]. In the case of certain parameter x, fuzzy number $\underline{A} = \{(x, \mu_{\underline{A}}(x))\}$ is the *singleton* with $x = x_{\min} = x_{\max}$ (more in [14]).

In our case the fuzzy approach replaces parameters I, NW, R_i, VC_i, FC_i, T_i with intervals $\langle I_{\min}, I_{\max}\rangle$, $\langle NW_{\min}, NW_{\max}\rangle$, $\langle R_{i\_min}, R_{i\_max}\rangle$, $\langle VC_{i\_min}, VC_{i\_max}\rangle$, $\langle FC_{i\_min}, FC_{i\_max}\rangle$, $\langle T_{i\_min}, T_{i\_max}\rangle$, i = 1, 2,…, n. By the application of operations of algebra of intervals defined in the article [7], the fuzzy approach replaces the equations $CF_0 = -(I + NW)$ and $CF_i = R_i - VC_i - FC_i - T_i$ with the couple of equations ($CF_{0min} = -(I_{max} + NW_{max})$; $CF_{0max} = -(I_{min} + NW_{min})$) and ($CF_{i\_min} = R_{i\_min} - VC_{i\_max} - FC_{i\_max} - T_{i\_max}$; $CF_{i\_max} = R_{i\_max} - VC_{i\_min} - FC_{i\_min} - T_{i\_min}$). The limits of intervals $\langle CF_{0min}, CF_{0max}\rangle$, $\langle CF_{i\_min}, CF_{i\_max}\rangle$ and $\langle IRR_{\min}, IRR_{\max}\rangle$ are inserted in the formula (1), the result of which is a pair of formulas (3):

$$\begin{aligned} \sum_{i=1}^{n} \mathrm{CF_{i\_min}} \,/(1 + IRR_{\min})^{\mathrm{i}} &= -\mathrm{CF_{0\_min}} \\ \sum_{i=1}^{n} \mathrm{CF_{i\_max}} \,/(1 + IRR_{\max})^{\mathrm{i}} &= -\mathrm{CF_{0\_max}} \end{aligned} \tag{3}$$

From formulas (3) the limits $IRR_{\min}$ and $IRR_{\max}$ of the output interval $\langle IRR_{\min}, IRR_{\max}\rangle$ can be calculated by any iterative method. The calculation of any value of the interval $\langle IRR_{\min}, IRR_{\max}\rangle$ is purely a technical matter that does not provide any reason for giving preference to one value over another. Consequently, the output interval $Y = \langle IRR_{\min}, IRR_{\max}\rangle$ is also the support of the fuzzy number $\underline{IRR} = \{(y, \mu_{\underline{IRR}}(y)): y \in \mathbb{R}\}$, $\mu_{\underline{IRR}}(y) = 1$ for $y \in Y$, $\mu_{\underline{IRR}}(y) = 0$ otherwise, therefore:

$$\begin{aligned} IRR &= \int_Y y \cdot \mu_{\underline{IRR}}(y) \cdot \mathrm{d}y \Big/ \int_Y \mu_{\underline{IRR}}(y) \cdot \mathrm{d}y = \\ &= (IRR_{\max}^2 - IRR_{\min}^2)/\big(2 \cdot (IRR_{\max} - IRR_{\min})\big) = (IRR_{\max} + IRR_{\min})/2 \end{aligned} \tag{4}$$

With regard to the fact that *IRR* calculated in this way is largely a consequence of the subjective experience and opinions of experts who define the intervals of possible values of uncertain parameters, we say that *IRR* is the *subjectively* expected value.

3 Application part: profitability analyses of electric car production in terms of uncertainty of data on the input side

The electric vehicle market is still at an early stage of development compared to the existing market of engine vehicles. Active government support is therefore a major factor in the creation and expansion of the electric car market. For example, in the United States the buyers and users of electric cars have received direct subsidies, tax relieves and/or tax exemption; in China, the government supports public vehicle purchases and a private electric car market in some pilot cities [15]. It is obvious that financial incentives are positively correlated with the market share of electric vehicles [12].

From the perspective of entrepreneurs, various reasons exist for considering the electric car production as a profitable business. In the countries following environmental regulations on greenhouse gas emissions, it is expected that the yield on such investments will be partially backed by public subsidies [13].

The illustration of how subsidies reflect in the profitability of investments in electric car production in terms of uncertainty of data on the input side is assessed according to IRR in the subsequently solved problem. The entrance data lean on the results of marketing research that have taken into account the actual situation regarding the subsidy policy in this field and its expected development in the next decade. Within it, we analyse a newly formed production of this commodity. The required minimum profitability of the investment r = 12 % reflects the requirement of shareholders and creditors investing their capital in this production and as such it is equal to the alternative market costs of capital in the automotive business area.

3.1 Task assignment

Consider a project of a seven-year cycle of production and sale of a selected number of electric vehicles of a given type whose basic long-term and short-term cost characteristics are recorded in rows 1 and 3 of column 0 and in rows 4 and 5 of Tab. 1, Tab. 2 and Tab. 3 in part 3.2. The company executive financial director highlights the difficulty in making accurate estimates of cash flow streams due to the fact that they deal with a new production and given to uncertain development of purchase subsidies for this commodity. His years´ of experience in the automotive business tells him that if the project is not to be loss-making, the value of the production margin x, defined by $x = (R - VC) / R$, where R is the revenue from sales and VC symbolizes the variable costs, has to reach at least 15 %. On the other hand, he also knows that it is almost impossible to exceed the 25 % limit given the current competition in the car market. This manager´s knowledge actually defines the interval ⟨15 %, 25 %⟩ of possible production margin values, within which he does not see any reason for giving preference to one value over another. The margin becomes the factor determining the limits of the interval of uncertain parameter R_i, for which it applies: $R_i = VC_i / (1 - x)$. Thus, $R_{i\_min} = VC_i / 0.85$, $R_{i\_max} = VC_i / 0.75$. The values R_{i_min}, or respectively, R_{i_max} are recorded in the second row of Tab. 2, or respectively, of Tab. 3.

3.2 Task solution

To analyse the impact of uncertainty of the parameter R_i of input variable CF_i, i = 1, 2,…, n on the budgeted cash flow we draw from Tab. 1, Tab. 2 and Tab. 3, in which input data of the baseline scenario are entered (about the cash flow budgeting in detail in [2] or [4]). In the second row of these tables the revenues from the sale of electro vehicles are estimated for the six years of the project run, the revenue in the 7th year stands for the income from the sale of dismantled machinery of a completely depreciated assembly line. Row 3 (working capital expenditure) presents the amount to be gradually put into operation from external sources during the first financial cycle. This amount remains „drowned“ in stocks, receivables and in product elaborations. It is gradually released at the shortening of the financial cycle and is fully released after the termination of production. In our case, the remainder of 1.5 mil. Euro will be dissolved in the non-profit production of spare parts. In row 8, the negative tax payment means a cash inflow resulting from a tax loss from a project, which is used to reduce the tax in other company production. From the tables it can be seen that the production gradually starts in the first year and reaches its full capacity in the 4th year, after which the monetary values of sales and costs only increase due to the expected two percent year-on-year inflation.

In Tab. 1 the conventional approach is solved, and therefore, the values given in row 2 are the centres of the intervals $\langle R_{i\_min}, R_{i\_max}\rangle$. By substituting the values of the row 11 into the formula (2) we get the equation: $14.34 / (1 + E[IRR]) + 20.7 / (1 + E[IRR])^2 + 22.33 / (1 + E[IRR])^3 + 24.04 / (1 + E[IRR])^4 + 24.5 / (1 + E[IRR])^5 + 25.3 / (1 + E[IRR])^6 + 5.28 / (1 + E[IRR])^7 = 81.6$, the solution of which is E[IRR] = 0.15, i.e., 15 %.

Year	0	1	2	3	4	5	6	7
1. Capital investment	75							
2. Revenue from sales		150	200	213.7	228.7	233.7	237.5	8
3. Production start-up investment	10						–8.5	
4. Variable costs		120	160	171	183	187	190	
5. Fixed costs		16	16.3	16.6	17	17.3	17.7	

6. Depreciation		15	15	15	15	15		
7. Profit before tax (2-3-4-5-6)	–10	–1	8.7	11.1	13.7	14.4	38.3	8
8. Tax 34 %	–3.4	–0.34	3	3.77	4.66	4.9	13	2.72
9. Net profit (7-8)	–6.6	–0.66	5.7	7.33	9.04	9.5	25.3	5.28
10. CF_i from operation (6+9)	–6.6	14.34	20.7	22.33	24.04	24.5	25.3	5.28
11. **E[CF_i]** (10-1)	**–81.6**	**14.34**	**20.7**	**22.33**	**24.04**	**24.5**	**25.3**	**5.28**

Table 1 Net cash flow forecast from the production and sale of electric vehicles in mil. EUR (conventional approach)

In Tab. 2 and Tab. 3, the fuzzy approach is solved and, therefore, the values given in row 2 of Tab. 2, or respectively, of Tab. 3, are the values of R_{i_min}, or respectively, R_{i_max}.

Year	0	1	2	3	4	5	6	7
1. Capital investment	75							
2. Revenue from sales (R_{i_min})		141.2	188.2	201.2	215.3	220	223.5	8
3. Production start-up investment	10						–8.5	
4. Variable costs		120	160	171	183	187	190	
5. Fixed costs		16	16.3	16.6	17	17.3	17.7	
6. Depreciation		15	15	15	15	15		
7. Profit before tax (2-3-4-5-6)	–10	–9.8	–3.1	–1.4	0.3	0.7	24.3	8
8. Tax 34 %	–3.4	–3.33	–1.05	–0.48	0.1	0.24	8.3	2.72
9. Net profit (7-8)	–6.6	–6.47	–2.05	–0.92	0.2	0.46	16	5.28
10. CF_{i_min} from operation (6+9)	–6.6	8.53	12.95	14.08	15.2	15.46	16	5.28
11. **CF_{i_min}** (10-1)	**–81.6**	**8.53**	**12.95**	**14.08**	**15.2**	**15.46**	**16**	**5.28**

Table 2 CF_{i_min} forecast from the production and sale of electric vehicles in mil. EUR (fuzzy approach)

Year	0	1	2	3	4	5	6	7
1. Capital investment	75							
2. Revenue from sales (R_{i_max})		160	213.3	228	244	249.3	253.3	8
3. Production start-up investment	10						–8.5	
4. Variable costs		120	160	171	183	187	190	
5. Fixed costs		16	16.3	16.6	17	17.3	17.7	
6. Depreciation		15	15	15	15	15		
7. Profit before tax (2-3-4-5-6)	–10	9	22	25.4	29	30	54.1	8
8. Tax 34 %	–3.4	3.06	7.48	8.64	9.86	10.2	18.4	2.72
9. Net profit (7-8)	–6.6	5.94	14.52	16.76	19.14	19.8	35.7	5.28
10. CF_{i_max} from operation (6+9)	–6.6	20.94	29.52	31.76	34.14	34.8	35.7	5.28
11. **CF_{i_max}** (10-1)	**–81.6**	**20.94**	**29.52**	**31.76**	**34.14**	**34.8**	**35.7**	**5.28**

Table 3 CF_{i_max} forecast from the production and sale of electric vehicles in mil. EUR (fuzzy approach)

Substituting the values of the row 11 of Tab. 2 and Tab. 3 into formulas (3) we get a pair of equations:

- $8.53 / (1 + IRR_{min}) + 12\ 95 / (1 + IRR_{min})^2 + 14.08 / (1 + IRR_{min})^3 + 15.2 / (1 + IRR_{min})^4 + 15.46 / (1 + IRR_{min})^5 + 16 / (1 + IRR_{min})^6 + 5.28 / (1 + IRR_{min})^7 = 81.6$ and
- $20.94 / (1 + IRR_{max}) + 29.52 / (1 + IRR_{max})^2 + 31.76 / (1 + IRR_{max})^3 + 34.14 / (1 + IRR_{max})^4 + 34.8 / (1 + IRR_{max})^5 + 35.7 / (1 + IRR_{max})^6 + 5.28 / (1 + IRR_{max})^7 = 81.6$,

the solution of which is $IRR_{min} = 0.02$, i.e., 2 %, $IRR_{max} = 0.28$, i.e., 28 % and the subjectively expected value $IRR = (2\ \% + 28\ \%) / 2 = 15\ \%$.

4 Discussion of the results and summary

The procedures and results of the study can briefly be summarized as follows:

- In this case, the statistically expected value E[IRR] of the project's return rate corresponds to its subjectively expected value of *IRR*; in general, this may not be the rule. The calculated value exceeds the required rate of return r = 12 % by 3 %, which supports the adoption of the project.
- The fuzzy approach brings additional information: possible yields can range from 2 % to 28 %. The probability that the required 12 % yield of return will not be reached is roughly equal to $100 \cdot 10 / 26 = 38$ %. In addition, the possibility also exists that the real yield of return is zero, as the 2 % yield of return is a nominal return, which would be absorbed by the expected 2 % year-on-year inflation rate. This can discourage an investor with a negative attitude to risk from investing.
- Furthermore, the fuzzy approach also informs that, for example, a 20 % capital investment subsidy that would reduce the initial investment from 75 to 60 mil. Euro would increase the minimum possible yield of return to 8 %. This could encourage investors to implement the investment. Thus, the electric vehicle market supply curve would shift to the right, i.e., in the same direction as the government-supported purchases (in order to increase consumption) shift the market demand curve.
- The degree of sensitivity of demand and supply to the government subsidies is different. The optimum is to divide the total incentive amount between the capital investment subsidy intended for the manufacturers and the subsidy directed to consumer purchases in such a way that the maximum level of consumption (purchases) would be achieved at the same price. The fuzzy approach has a potential to make a significant contribution to this topic.

References

[1] Běhounek, L., & Cintula, P. (2006). From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, ***157***(5), 642-646.

[2] Brealey, R. A., Myers, S. C. & Marcus, A. J. (2011). *Fundamentals of Corporate Finance*. Mcgraw-Hill Education, New York.

[3] Danielson, M. G. (2016). The IRR of a project with many potential outcomes. *The Engineering Economist*, ***61***(1), 44-56.

[4] Elkjaer, M. (2000). Stochastic budget simulation. *International Journal of Project Management*, ***18***(2), 139-147.

[5] Fiala, P. & Borovička, A. (2015). Investment decision-making by a two-step multi-criteria procedure. *International Journal of Multicriteria Decision Making*, ***5***(3), 274-293.

[6] Hašek, R. (2012). Numerical analysis of a planar motion; GeoGebra as a tool of investigation. *North American GeoGebra Journal*, ***1***(1).

[7] Hašková, S. (2017). Contribution to Economic Efficiency Evaluation of Projects in Terms of Uncertainty. In Prazak, P. *35th International Conference Mathematical Methods in Economics* (*MME*). 1. vyd. Hradec Králové, ČR: University of Hradec Králové, 213-218.

[8] Maciel, L., Gomide, F., & Ballini, R. (2016). Evolving fuzzy-GARCH approach for financial volatility modeling and forecasting. Computational Economics, ***48***(3), 379-398.

[9] Noor, M. A., & Waseem, M. (2009). Some iterative methods for solving a system of nonlinear equations. *Computers & Mathematics with Applications*, ***57***(1), 101-106.

[10] Patrick, M., & French, N. (2016). The internal rate of return (IRR): projections, benchmarks and pitfalls. *Journal of Property Investment & Finance*, ***34***(6), 664-669.

[11] Pettigrew, R. (2014). Accuracy, Risk, and the Principle of Indifference (forthcoming in Philosophy and Phenomenological Research). *Philosophy and Phenomenological Research*, ***92***(1), 35-59.

[12] Sierzchula, W., Bakker, S., Maat, K., & Van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy*, ***68***, 183-194.

[13] Yong, T., & Park, C. (2017). A qualitative comparative analysis on factors affecting the deployment of electric vehicles. *Energy Procedia*, ***128***, 497-503.

[14] Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, ***30***(3-4), 407-428.

[15] Zhang, X., Xie, J., Rao, R., & Liang, Y. (2014). Policy incentives for the adoption of electric vehicles across countries. *Sustainability*, ***6***(11), 8056-8078.

Identification of the Maximal Relevant Distance in Emergency System Designing

Marek Kvet[1], Jaroslav Janáček[2]

Abstract. Emergency service system structure is determined by deployment of given number of service center locations. The objective is to minimize the average response time of system users. Advanced exact methods for obtaining the optimal deployment of the service centers are based on exploitation of a radial model of the location problem. The efficiency of this approach is strongly influenced by the number of considered zones, which is proportional to the maximal relevant distance. The maximal relevant distance in the solved location problem is the maximal distance between user and possible service center locations, which must be taken into account unless the solving process losses the optimal solution. The paper deals with the methods of the maximal relevant distance determination in the location problems, where the system disutility of a user is proportional to the distance from the user location to the nearest located service center. We also enrich the paper with a study of the problems, in which the generalized disutility is considered. The generalized disutility of a user takes into account distances to more than one nearest located service centers.

Keywords: Emergency service system, radial approach, maximal relevant distance

JEL Classification: C44
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

This paper deals with Emergency Medical Service (EMS) system designing. Such and similar systems are designed for given geographical area to satisfy the demand of population living in the area for more secure life. The associate service is provided from a given number of service centers and the overwhelming objective used for the design evaluation is to minimize the total discomfort of all system users measured by the average time necessary to deliver service from a center to the user location, at which the demand for service has occurred [1], [2], [5], [10].

Host of models consider that the serviced population is concentrated to a finite number of dwelling places of the considered area. Frequency of the demand occurrence is proportional to the number of inhabitants of the given town or village. A finite set of possible service center locations is assumed and also, the assumption is made that a user demand is serviced from the nearest located service center. This way, the weighted p-median problem formulation is used to the emergency service system design and solving the underlying problem to optimality [1], [2], [3], [7]. If a large instance of the problem is described by a location-allocation model, then the model size often exceeds any acceptable limit for available optimization software. It must be noted, that the numbers of user locations and possible service center locations may take the value of several hundreds or thousands [4]. Furthermore, the number of possible center locations seriously impacts the computational time and memory demands due to storing the unfathomed nodes of the inspected searching tree for further processing by the branch-and-bound method. That is why the direct attempt at solving the problem described by a location-allocation model often fails. To avoid this obstacle, the approximate approach based on a radial formulation with homogenous system of radii has been developed [4], [8], [9]. This approach uses an approximation of a common time distance between a service center location and a user location by some pre-determined time distances and gives either a near-to-optimal result in the case of integer time distances or the exact one, if all possible values from the matrix are taken into account. As the radial formulation of the problem avoids assigning a center to a user like it is common in the location-allocation approach, the used model is smaller than the location-allocation one. In addition, the solving method used in the IP-solvers converges much faster. On the other hand, it can be observed that the efficiency of suggested radial approach is strongly influenced by the number of considered zones, which is proportional to the maximal relevant distance [4]. The maximal relevant distance in the solved location problem is the maximal distance between user and possible service center locations, which must be taken into account unless the solving

[1] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, marek.kvet@fri.uniza.sk
[2] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, jaroslav.janacek@fri.uniza.sk

process losses the optimal solution. Our main research goal consists in identification of the maximal relevant distance, which enables to achieve considerable computational time reduction without any loss of solution accuracy. In addition, we perform the computational study with an extended version of the associated location problem. The mentioned extension consists in incorporating the concept of so-called generalized disutility into the mathematical model. It enables us to consider the stochastic feature of the modelled system in a partial content.

The concept of generalized disutility introduced in [7], [8] follows the idea that the service does not need to be necessarily delivered only from the nearest located service center. Thus, the generalized disutility of any user takes into account distances to more than one nearest located service centers. At the moment of current demand occurrence, the nearest service center may be temporarily unavailable due to satisfying another demand. Therefore, the original objective function minimizing the average time distance to the nearest located service center from a user can be replaced by minimization of the average time distance of a user to the expected nearest available center.

2 Radial Model of EMS Design Problem with Generalized Disutility

To formulate the radial model of EMS design problem, several notations must be introduced. Let the symbol J denote a finite set of all user locations (dwelling places), where b_j denotes the number of users sharing the location $j \in J$. Let I be a finite set of all possible service center locations, from which exactly p centers should be chosen. We use the symbol d_{ij} to denote the integer network time distance between locations i and j, where $i \in I$ and $j \in J$. The maximal considered distance is m, i.e. $m = \max\{d_{ij}: i \in I, j \in J\}$.

Let the value of q_k represent the probability of the case that the k-1 nearest centers are temporarily unavailable and the k-th nearest center is the first one, which can provide the demanded service to the user [6], [8]. The probabilities q_k for $k = 1, ..., r$ are positive real values, which meet the following inequalities $q_1 \geq q_2 \geq \ldots \geq q_r$. In this paper, the generalized disutility perceived by a user is modelled by a sum of weighted time distances from the r nearest located centers. To complete the radial model, we introduce coefficients a^s_{ij} for each pair i, j of possible center location and a user location. We define $a^s_{ij} = 1$ if $d_{ij} \leq s$ and $a^s_{ij} = 0$ otherwise for each $s = 0, 1, \ldots, m$-1. Finally, let us introduce the remaining decision variables. To describe decisions on the resulting service center deployment, we introduce zero-one variable y_i for each $i \in I$. The variable y_i takes the value of 1, if a service center is to be located at i and it takes the value of 0 otherwise. To express the total distance necessary for user demand satisfaction in the radial manner, we introduce auxiliary zero-one variables x_{jsk} for $j \in J$, $s \in 0, ..., m$-1, $k \in 1, ..., r$ to model the disutility contribution of the k-th nearest service center to the user located at j. The variable x_{jsk} takes the value of 1 if the k-th smallest disutility contribution for the user located at $j \in J$ is greater than s and it takes the value of 0 otherwise. Then the expression $x_{j0k} + x_{j1k} + \ldots + x_{j(m-1)k}$ constitutes the k-th smallest time distance from the user j to a located center. Using the above introduced assumptions, we suggest the following model [8].

$$\textit{Minimize} \qquad \sum_{j \in J} b_j \sum_{s=0}^{m-1} \sum_{k=1}^{r} q_k x_{jsk} \tag{1}$$

$$\textit{Subject to:} \quad \sum_{k=1}^{r} x_{jsk} + \sum_{i \in I} a^s_{ij} y_i \geq r \qquad \textit{for } j \in J,\ s = 0,\ldots, m-1 \tag{2}$$

$$\sum_{i \in I} y_i = p \tag{3}$$

$$y_i \in \{0,1\} \qquad \textit{for } i \in I \tag{4}$$

$$x_{jsk} \in \{0,1\} \qquad \textit{for } j \in J,\ s = 0,\ldots, m-1,\ k = 1,\ldots, r \tag{5}$$

The objective function (1) expresses the expected volume of generalized disutility. For given pair of user location j and a distance value s, the constraint (2) expresses relation between the set of location variables y_i, $i \in I$ and the sum of auxiliary variables x_{jsk} over the range 1, …, r of subscript k. If no center is located in the radius s, then the sum of auxiliary variables x_{jsk} equals to r. If exactly $k \leq r$ centers are located in the radius s, then the sum of variables equals to r-k due to the minimization process, which presses down the values of the variables x_{jsk}. If the sum of variables x_{jsk} equals to $k < r$, then the variables $x_{js1}, \ldots, x_{js(r-k)}$, equal to 0 and remaining variables equal to 1 due to the used optimization process and decreasing values of the coefficients $q_1, \ldots, q_r$. Finally, the constraint (3) limits the number of located service centers.

Even if the problem in the radial form is smaller and easier to solve than the same instance described by the location-allocation formulation, memory insufficiency may occur, when large instance of the weighted p-median problem is solved using commercial IP-solver. Such an instance is usually characterized by considerably big value of m, which means either too many different values in the distance matrix or too high maximal time distance value. In such case, the presented exact approach (1) - (5) can be reduced by proper selection of distance values, which should be taken into account. We assume that the higher distance values do not represent the time distances from user location to their r nearest located service centers. If we were able to identify the maximal relevant distance v occurring in the unknown optimal solution, we could replace the range 0, …, m-1 by the range 0, …, v-1, where $v << m$, and solve a smaller model without any loss of solution accuracy. The goal of making the model size smaller leads also to considerable computational time reduction.

Determination of suitable value of v is not easy, because of two contradictory demands. The first demand consists in maximal accuracy of the obtained solution. This demand may issue into very large size of associated model. The second demand asks for acceptable computational time and tries to minimize the model size as much as possible. Therefore, we deal here with a proper way of identification of the maximal relevant distance, which enables to comply with both demands to certain extent.

3 Maximal Relevant Distance Determination

This section is devoted to the maximal relevant distance determination. As we have discussed in the previous part, the maximal relevant distance may significantly reduce the computational time and the memory demands of the associated solving process, when the exact solution of the model (1) - (5) is being searched.

The maximal relevant distance can be determined by solving the following p-center problem (6) - (12), which uses the same notations as the model before. Since we use here only the basic version of the model without the concept of generalized disutility, the variables x_{js} do not need to take the index k as before.

$$Minimize \quad h \tag{6}$$

$$Subject\ to: \quad x_{js} + \sum_{i \in I} a_{ij}^{s} y_i \geq 1 \qquad for\ j \in J,\ s = 0, ..., m-1 \tag{7}$$

$$\sum_{i \in I} y_i = p \tag{8}$$

$$\sum_{s=0}^{m-1} x_{js} \leq h \qquad for\ j \in J \tag{9}$$

$$y_i \in \{0,1\} \qquad for\ i \in I \tag{10}$$

$$x_{js} \in \{0,1\} \qquad for\ j \in J,\ s = 0, ..., m-1 \tag{11}$$

$$h \geq 0 \tag{12}$$

The objective function (6) represented by a single variable h gives the upper bound of all perceived disutility values. The constraints (7) ensure that the variables x_{js} are allowed to take the value 0, if there is at least one service center located in the radius s from the user location j. The constraint (8) puts the limit p on the number of located service centers. Finally, the constraints (9) ensure that each perceived disutility is less than or equal to h.

The presented model (6) - (12) is hard to be solved mainly due to the min-max objective (6) and the link-up constraints (9). Fortunately, it can be solved exactly very fast by a bisection radial approach [8]. The key idea consists in iterative processing of the range 0, ..., m and searching for the smallest value of v from this range, for which all users are closer to their nearest centers than the value of s. In each iteration, a simple covering problem is solved. After the optimal solution of the model (6) - (12) is obtained, the resulting value of v represents the maximal relevant distance for the model (1) - (5). Nevertheless, determination of the proper value of the maximal relevant distance asks for a more sophisticated approach, which is based on the following analysis. The problem (1) - (5) can be formulated in the combinatorial way by (13).

$$F^{opt} = \min\left\{\sum_{j\in J} b_j \sum_{k=1}^{r} q_k \min\left\{d_{ij} \ : i \in I_1\right\} : I_1 \subset I, |I_1| = p\right\} \tag{13}$$

The $d_{ij} < m$ for $i \in I$, $j \in J$ and $\min_k\{d_{ij}: i \in I_1\}$ denotes the k-th minimal value of the set $\{d_{ij}: i \in I_1\}$ of distances for a given user j. Let us formulate the truncated problem for a given value v by (14).

$$F^{opt}(v) = \min\left\{\sum_{j\in J} b_j \sum_{k=1}^{r} q_k \min_k\left\{\min\left\{d_{ij}, v\right\} : i \in I_1\right\} : I_1 \subset I, |I_1| = p\right\} \tag{14}$$

Obviously $F^{opt}(m) = F^{opt}$. Further, we prove the following theorem.

Theorem 1: If the optimal solution $I_1{}^{opt}(v)$ of (14) exists and the expression (15) holds, then $I_1{}^{opt}(v)$ is the optimal solution of (13).

$$\min_k\left\{\min\left\{d_{ij}, v\right\} : i \in I_1^{opt}(v)\right\} < v \qquad for\ j \in J, k = 1,..., r \tag{15}$$

Proof: The inequality (16) follows from optimality $I_1{}^{opt}(v)$.

$$\begin{aligned} F^{opt}(v) = \sum_{j\in J} b_j \sum_{k=1}^{r} q_k \min_k\left\{\min\left\{d_{ij}, v\right\} : i \in I_1^{opt}(v)\right\} \leq \\ \leq \min\left\{\sum_{j\in J} b_j \sum_{k=1}^{r} q_k \min_k\left\{\min\left\{d_{ij}, v\right\} : i \in I_1\right\} : I_1 \subset I, |I_1| = p\right\} \end{aligned} \tag{16}$$

As $\min_k\{\min\{d_{ij}: v\}: i \in I_1\} \leq \min_k\{d_{ij}: i \in I_1\}$ holds for each $I_1 \subset I$, then $F^{opt}(v) \leq F^{opt}$. Then, we can obtain (17) from the expression (15).

$$F^{opt}(v) = \sum_{j\in J} b_j \sum_{k=1}^{r} q_k \min_k\left\{\min\left\{d_{ij}, v\right\} : i \in I_1^{opt}(v)\right\} = \sum_{j\in J} b_j \sum_{k=1}^{r} q_r \min_k\left\{d_{ij} \ : i \in I_1^{opt}(v)\right\} = F^{opt} \tag{17}$$

The theorem 1 can be easily used for size reduction of the problem (1) - (5). The process of reduction can start with some estimated relevant value v, which is considerably less than the original maximal distance m. Then, the model (1) - (5) is reformulated for the maximal distance $\underline{m} = v$ and the reduced problem is solved to optimality. The optimal solution is inspected concerning the variables $x_{j,\,m-1,\,k}$ for $j \in J$ and $k = 1, \ldots, r$. If all inspected variables equal to zero, then constraints (15) are satisfied and the obtained solution of the reduced problem is also the optimal solution of the original problem. In such a case, when some of the inspected variables equals to one, the sampling relevant value v must be incremented by a value d and the reduction process continues with the next trial until constraints (15) are met. Efficiency of the reduction process obviously depends on suitable choice of the starting value v and the increment d. The suitable setting of the two values will be studied in the next section.

4 Computational Study

The main goal of this computational study is to develop and verify an effective method for solving the EMS design problem with generalized disutility described by the model (1) - (5) associated with the min-max model (6) - (12), which helps us not only to determine the maximal relevant distance, but using this value enables us to reduce the size of the original problem and directly to decrease the computational time and memory demands. All numerical experiments presented in this section were performed in the optimization software FICO Xpress 7.3, 64-bit. The experiments were run on a PC equipped with the Intel® Core™ i7 5500U 2.4 GHz processor and 16 GB RAM.

Used problem instances were derived from real emergency health care system, which was implemented in eight regions of Slovakia separately. To enrich the set of benchmarks, we have added one instance, in which the optimal deployment of service centers for the whole Slovak Republic was searched. For each self-governing region of Slovakia, i.e. Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA), all cities and villages with corresponding number of inhabitants b_j were taken into account. The coefficients b_j were rounded up to hundreds. The set of communities represents both the set J of users' locations and the set I of possible service center locations as well. The cardinalities of these sets vary from 87 to 664 locations. An individual experiment was organized so that the model (1) - (5) was solved for $r = 3$. As

discussed besides in [9], three nearest service centers are enough to model real emergency medical service system with satisfactory solution accuracy. The associated coefficients q_k for k=1, …, r were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. These values were obtained from a simulation model of existing EMS system in Slovakia as described in [6].

The following Table 1 summarizes the preliminary results obtained for two selected problem instances solved for the self-governing regions Žilina ($|I| = 315$, $p = 32$) and Prešov ($|I| = 664$, $p = 67$) respectively. Each row of the table corresponds to one setting of v, for which the radial model (1) - (5) was solved. The initial value of v was obtained by the min-max model (6) - (12). The computational process searching for the optimal solution of the model (6) - (12) takes approximately 0.2 seconds for Žilina and it does not exceed 1.3 seconds for Prešov. Each obtained solution of the problem described by terms (1) - (5) is described by three different values: The computational time in seconds is reported in columns denoted by "*CT* [s]". Since the radial model (1) - (5) with $v < m$ approximates all distances, which are higher than v by the value of v, the objective function of the resulting solution may be lower than the real objective function computed for the obtained solution based on the distance matrix $\{d_{ij}\}$. As a consequence of this approximation, the computational process may lose the optimal solution of the original problem, if the used value of v does not cover all relevant distances, which occur in the resulting solution. If the objective functions equal, than we know that the optimal solution of the original problem is obtained. Therefore, we report both objective function values. The model objective function value is reported in columns denoted by "Model ObjF" and the associated real objective value is denoted by "realObjF". The last row of the table corresponds to such problem instance, in which the full set of distances was used, i.e. $v = m$.

v	Žilina ($\|I\| = 315$, $p = 32$)			Prešov ($\|I\| = 664$, $p = 67$)		
	CT [s]	Model ObjF	realObjF	*CT* [s]	Model ObjF	realObjF
14	0.14	32898.17	35311.11	0.41	31511.14	33346.68
16	0.19	33743.58	35080.76	0.67	32032.29	33247.73
18	0.28	34314.47	35080.76	0.80	32381.58	33057.80
20	0.28	34605.46	34933.30	1.22	32582.01	33050.86
22	0.47	34771.87	34933.30	1.14	32725.15	33050.86
24	0.39	34851.74	34933.30	1.34	32806.19	32919.24
26	0.47	34889.21	34933.30	1.53	32840.40	32899.29
28	0.50	34913.12	34933.30	1.81	32863.79	32899.29
30	0.56	34921.88	34933.30	2.00	32878.10	32899.29
32	0.66	34927.88	34933.30	2.36	32887.62	32899.29
34	0.66	34930.97	34933.30	2.88	32892.98	32899.29
36	0.74	34932.20	34933.30	3.00	32896.06	32899.29
38	0.78	34932.91	34933.30	3.63	32897.81	32899.29
40	0.88	34933.17	34933.30	4.16	32899.10	32899.29
42	1.02	34933.30	34933.30	4.89	32899.29	32899.29
44	1.08	34933.30	34933.30	5.20	32899.29	32899.29
46	1.14	34933.30	34933.30	6.28	32899.29	32899.29
48	1.20	34933.30	34933.30	6.56	32899.29	32899.29
50	1.36	34933.30	34933.30	12.01	32899.29	32899.29
52	1.38	34933.30	34933.30	14.60	32899.29	32899.29
54	1.53	34933.30	34933.30	16.24	32899.29	32899.29
m	5.91	34933.30	34933.30	125.25	32899.29	32899.29

Table 1 Impact of the maximal relevant distance on the model (1) - (5) and the resulting EMS design

The results in Table 1 show, that the optimal maximal relevant distance is around 40. If we concentrate on the computational time, the determination of the maximal relevant distance enables us to obtain the optimal solution of the original problem much faster. The comparison of the min-max solution (initial value of v), original value of m and the optimal maximal relevant distance for all solved problem instances is reported in the following Table 2. The initial value of v is denoted by "v^{init}" and the optimal maximal relevant distance is denoted by "v^{opt}".

When looking at the reported results, it must be noted, that the mathematical model (1) - (5) with the full distance set, i.e. v=m, computed for the whole Slovak Republic ($|I| = 2916$, $p = 273$) is not solvable due to the model size and very high memory demands - the model contains more than 2916*546*3 variables. Nevertheless, the optimal solution can be obtained making use of $v = 46$ within one minute. Thus, the determination of the maximal relevant distance brings solvability to such problem instances, which would not be solvable otherwise.

Based on performed numerical experiments (see Table 2) and according to the analysis of the impact of the maximal relevant distance on the computational time, we suggest the following algorithm: Solve the problem (6) - (12) to obtain the initial value of v^{init}. Then, set v^{opt} to the value of $4*v^{init}$ and solve the model (1) - (5) for v^{opt}. If not all relevant distance values are covered by v^{opt}, set $v^{opt} = 5*v^{init}$ and solve the model (1) - (5) again. Thus, we suggest the value d in the introduced algorithm to be equal to v^{init}.

region	$\|I\|$	p	m	v^{init}	v^{opt}
BA	87	9	79	14	46
BB	515	52	182	13	30
KE	460	46	200	12	42
NR	350	35	133	13	42
PO	664	67	253	12	41
TN	276	28	149	12	42
TT	249	25	157	13	39
ZA	315	32	155	14	41
SR	2916	273	546	13	46

Table 2 Maximal relevant distances for solved problem instances

5 Conclusions

The paper offers the methods of reducing the size of the Emergency Medical Service system design problem by determination of the maximal relevant distance. This reduction enables considerable diminishing of the associated model and acceleration of the optimization process. Professionals in the field of emergency system designing could appreciate the results of the performed computational study, which issues in the suggested rule for preliminary maximal relevant distance determination.

Acknowledgements

This work was supported by the research grants VEGA 1/0342/18 "Optimal dimensioning of service systems", VEGA 1/0089/19 "Data analysis methods and decisions support tools for service systems supporting electric vehicles", and VEGA 1/0689/19 "Optimal design and economically efficient charging infrastructure deployment for electric buses in public transportation of smart cities" and APVV-15-0179 "Reliability of emergency systems on infrastructure with uncertain functionality of critical elements".

References

[1] Brotcorne, L. Laporte, G. & Semet, F. (2003). Ambulance location and relocation models. In *European Journal of Operational Research* 147, pp. 451-463.

[2] Correia, I. & Saldanha da Gama, F. (2015). Facility locations under uncertainty. *Location Science*, eds. Laporte, Nikel, Saldanha da Gama, pp. 177-203.

[3] Czimmermann, P. & Kohání, M. (2018). Computation of transportation performance in public service systems. In *IEEE workshop on complexity in engineering*, Firenze, pp. 1-5.

[4] García, S. Labbé, M. & Marín, A. (2011). Solving large p-median problems with a radius formulation. In *INFORMS Journal on Computing*, Vol. 23, No 4, 546-556.

[5] Ingolfsson, A., Budge, S. & Erkut, E. (2008). Optimal ambulance location with random delays and travel times, In *Health Care Management Science*, Vol. 11, No 3, pp. 262-274.

[6] Jankovič P. (2016). Calculating Reduction Coefficients for Optimization of Emergency Service System Using Microscopic Simulation Model. In *17th International Symposium on Computational Intelligence and Informatics*, Budapest, Hungary, pp. 163-167.

[7] Jánošíková, Ľ. & Žarnay, M. (2014). Location of emergency stations as the capacitated p-median problem. In *Quantitative Methods in Economics-Multiple Criteria Decision Making XVII*, Virt, Slovakia.

[8] Kvet, M. (2014). Computational Study of Radial Approach to Public Service System Design with Generalized Utility. In *Digital Technologies 2014*, Žilina, Slovakia, pp. 198-208.

[9] Kvet, M. & Janáček, J. (2018). Reengineering of the Emergency Service System under Generalized Disutility. In *ICORES 2018*, Madeira, Portugal, pp. 85-93.

[10] Marianov, V. & Serra, D. (2002). *Location problems in the public sector*. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp. 119-150.

Tolerance Solvability of Interval Max-min Matrix Equations

Helena Myšková[1]

Abstract. Behavior of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of linear equations or by matrix equations in max-min algebra. Max-min algebra is an algebraic structure, in which classical addition and multiplication are replaced by maximum and minimum, respectively. Max-min equations have found a broad area of applications in causal models which emphasize relationships between input and output variables. They are used in diagnosis models or models of non-deterministic systems. Many practical situations can be described using max-min matrix equations. It often happens that a max-min matrix equation with exact data is unsolvable. Therefore, we replace matrix elements with intervals of possible values. In this way, we obtain an interval matrix equation. We can define several types of solvability of interval max-min matrix equation.
In this paper, we shall deal with the possibly tolerance and strongly tolerance solvability of interval max-min matrix equations. We prove the necessary and sufficient conditions which can be checked in polynomial time.

Keywords: max-min algebra, interval matrix, matrix equation, strongly tolerance solvability, weakly tolerance solvability

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Introduction

Behavior of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of linear equations or by matrix equations. Discrete dynamic systems and related algebraic structures were studied using max-min matrix operations in [1], [3]. In the last decades, significant effort has been developed to study systems of max-min linear equations in the form $A \otimes x = b$, where A is a matrix, b and x are vectors of compatible dimensions. Systems of linear equations over max-min algebra are used in several branches of applied mathematics. Among interesting real-life applications let us mention e.g. a large scale model of Dutch railway network or synchronizing traffic lights in Delfts [14]. In the last two decades, interval systems of the form $\boldsymbol{A} \otimes x = \boldsymbol{b}$ have been studied, for details see [4] –[8].

In this paper, we shall deal with matrix equations of the form $A \otimes X \otimes C = B$, where $A,\ B$, and C are given matrices of suitable sizes a X is an unknown matrix. The following example is the shortened version of an example given in [2].

Example 1. Let us consider a situation, in which passengers from places $P_1, P_2, \ldots, P_m$ want to transfer to holiday destinations $D_1, D_2, \ldots, D_r$. Different transportation means provide transporting passengers from places $P_1, P_2, \ldots, P_m$ to airport terminals $T_1, T_2, \ldots, T_s$. We assume that the connection between P_i and T_l is possible only via one of the check points $Q_1, Q_2, \ldots, Q_n$.

Denote by a_{ij} (c_{lk}) the known capacity of the connection between P_i and Q_j (T_l and D_k). If there is no connection from P_i to Q_j (between T_l and D_k), we put $a_{ij} = 0$ ($c_{lk} = 0$). If Q_j is linked with T_l by a road with the capacity x_{jl}, then the capacity of the connection between P_i and D_k via Q_j using terminal T_l is equal to $\min\{a_{ij}, x_{jl}, c_{lk}\}$.

Suppose that there are b_{ik} passengers travelling from place P_i to destination D_k. Our task is to choose the appropriate capacities x_{jl} for any $j \in N, N = \{1,\ 2, \ldots,\ n\}$ and for any $l \in S,\ S = \{1, 2, \ldots, s\}$ such that the maximum capacity of the road from P_i to D_k is equal to b_{ik} for any $i \in M, M = \{1, 2, \ldots, m\}$ and for any $k \in R,\ R = \{1, 2, \ldots, r\}$, i. e.,

$$\max_{j \in N,\, l \in S} \min\{a_{ij}, x_{jl}, c_{lk}\} = b_{ik} \tag{1}$$

has to be satisfied.

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, helena.myskova@tuke.sk

2 Preliminaries

Max-min algebra is the triple $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I} = [O, I]$ is a linear ordered set with the least element O and the greatest element I and $\oplus$, $\otimes$ are binary operations defined as follows:

$$a \oplus b = \max\{a, b\} \text{ and } a \otimes b = \min\{a, b\}.$$

The set of all $m \times n$ *matrices* over $\mathcal{I}$ is denoted by $\mathcal{I}(m, n)$ and the set of all column n-*vectors* over $\mathcal{I}$ by $\mathcal{I}(n)$.
Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in the classical algebra. We consider the *ordering* $\leq$ on the sets $\mathcal{I}(m, n)$ and $\mathcal{I}(n)$ defined as follows:

- for $A, C \in \mathcal{I}(m, n)$: $A \leq C$ if $a_{ij} \leq c_{ij}$ for each $i \in M$ and for each $j \in N$,
- for $x, y \in \mathcal{I}(n)$: $x \leq y$ if $x_j \leq y_j$ for each $j \in N$.

We will use the *monotonicity of* $\otimes$, which means that for each $A, C \in \mathcal{I}(m, n)$ and for each $B,\ D \in \mathcal{I}(n, s)$ the implication

$$\text{if } A \leq C \text{ and } B \leq D \text{ then } A \otimes B \leq C \otimes D$$

holds. Let $A \in \mathcal{I}(m, n)$ and $b \in \mathcal{I}(m)$. We can write the system of max-min linear equations in the matrix form

$$A \otimes x = b. \tag{2}$$

It is well known (see [1, 16]) that system (2) is solvable if and only if the vector $x^*(A, b)$, defined by

$$x_j^*(A, b) = \min_{i \in M}\{b_i : a_{ij} > b_i\} \tag{3}$$

for any $j \in N$, where $\min \emptyset = I$, is its solution. The vector $x^*(A, b)$ is called a *principal solution* of system (2).

This article deals with matrix equations of the form

$$A \otimes X \otimes C = B \tag{4}$$

where $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(s, r)$ are given matrices and we are looking for $X \in \mathcal{I}(n, s)$.

Returning to Example 1, equality (1) can be written in max-min algebra in the form $[A \otimes X \otimes C]_{ik} = b_{ik}$. Therefore, the solvability of the problem in mentioned example is equivalent to the solvability of (4). In [2], there were defined so-called *principal matrix solution* $X^*(A, B, C) \in \mathcal{I}(n, s)$, defined as follows:

$$x_{jl}^*(A, B, C) = \min_{k \in R} x_j^*(A \otimes c_{lk}, B_k) \tag{5}$$

for each $j \in N$ and for each $l \in S$. The following theorem expresses the importance of the principal matrix solution for the solvability of (4).

Theorem 1 (Draženská–Myšková [2]). *Let $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(m, n)$.*
(i) *If $A \otimes X \otimes C = B$ for $X \in \mathcal{I}(n, s)$, then $X \leq X^*(A, B, C)$;*
(ii) $A \otimes X^*(A, B, C) \otimes C \leq B$;
(iii)*The matrix equation $A \otimes X \otimes C = B$ is solvable if and only if the matrix $X^*(A, B, C)$ is its solution.*

The following lemma describes the properties of a principal matrix solution.

Lemma 2 (Draženská–Myšková [2]). Let $A,\ A^{(1)},\ A^{(2)} \in \mathcal{I}(m, n)$, $B,\ B^{(1)},\ B^{(2)} \in \mathcal{I}(m, r)$ and $C,\ C^{(1)},\ C^{(2)} \in \mathcal{I}(s, r)$. The following assertions hold:
(i) if $A^{(2)} \leq A^{(1)}$ then $X^*(A^{(1)}, B, C) \leq X^*(A^{(2)}, B, C)$;
(ii) if $B^{(1)} \leq B^{(2)}$ then $X^*(A, B^{(1)}, C) \leq X^*(A, B^{(2)}, C)$;
(iii)if $C^{(2)} \leq C^{(1)}$ then $X^*(A, B, C^{(1)}) \leq X^*(A, B, C^{(2)})$.

3 Interval Matrix Equations

A certain disadvantage of the necessary and sufficient condition for the solvability of (4) given in Theorem 1 (iii) stems from the fact that it only indicates the existence or non-existence of the solution but does not indicate any action to be taken to increase the degree of solvability. However, it happens quite often in modelling real situations that the obtained system turns out to be unsolvable.

One of possible methods of restoring the solvability is to replace the exact input values by intervals of possible values. In practice, the capacities a_{ij}, c_{lk} in Example 1 may depend on external conditions, so they are from intervals of possible values. Due to this fact, we will require the capacities of connections to be from a given interval of possible values.

Similarly to [9, 10, 11, 15], we define *interval matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ as follows:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{I}(m,n);\ \underline{A} \leq A \leq \overline{A} \right\},$$

$$\boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \mathcal{I}(m,r);\ \underline{B} \leq B \leq \overline{B} \right\},$$

and

$$\boldsymbol{C} = [\underline{C}, \overline{C}] = \left\{ C \in \mathcal{I}(s,r);\ \underline{C} \leq C \leq \overline{C} \right\}.$$

Denote by

$$\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B} \tag{6}$$

the set of all matrix equations of the form (4) such that $A \in \boldsymbol{A}$, $B \in \boldsymbol{B}$ and $C \in \boldsymbol{C}$. We call equation (6) an *interval max-min matrix equation*.

We shall think over the solvability of interval max-min matrix equation on the ground of solvability of matrix equations of the form (4) where $A \in \boldsymbol{A}$, $B \in \boldsymbol{B}$, and $C \in \boldsymbol{C}$. We can define several types of solvability of interval max-min matrix equations. Some of them have been studied in [2]. In the following, we define other five solvability concepts.

3.1 Strongly universal, universal and weak universal solvability

Definition 1. Interval max-min matrix equation (6) is strongly universally solvable if there exists a matrix $X \in \mathcal{I}(n,s)$ such that the equality $A \otimes X \otimes C = B$ is satisfied for each $A \in \boldsymbol{A}$, for each $C \in \boldsymbol{C}$ and for each $B \in \boldsymbol{B}$.

Theorem 3 (Myšková–Plavka [13])**.** *Interval max-min matrix equation* (6) *is strongly universally solvable if and only if* $\underline{B} = \overline{B} = B$ *and*

$$\underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C} = B. \tag{7}$$

Definition 2. Interval max-min matrix equation (6) is universally solvable if for each $B \in \boldsymbol{B}$ there exists a matrix $X \in \mathcal{I}(n,r)$ such that the equality $A \otimes X \otimes C = B$ holds for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$.

To give an equivalent condition for the universal solvability let us define for each $p \in M$ and for each $u \in R$ the matrix $B^{(pu)}$ as follows:

$$b_{ik}^{(pu)} = \begin{cases} \overline{b}_{ik} & \text{for } (i,k) = (p,u), \\ \underline{b}_{ik} & \text{otherwise.} \end{cases} \tag{8}$$

Theorem 4 (Myšková–Plavka [13])**.** *Interval max-min matrix equation* (6) *is universally solvable if and only if*

$$\underline{A} \otimes X^*(\overline{A}, B^{(pu)}, \overline{C}) \otimes \underline{C} = B^{(pu)} \tag{9}$$

for each $p \in M$ *and for each* $u \in R$.

Definition 3. Interval max-min matrix equation (6) is weakly universally solvable if a max-min matrix equation $A \otimes X \otimes C = B$ is solvable for each $A \in \boldsymbol{A}$, for each $B \in \boldsymbol{B}$, and for each $C \in \boldsymbol{C}$.

Theorem 5 (Myšková–Plavka [13])**.** *Interval max-min matrix equation* (6) *is weakly universally solvable if and only if it is universally solvable.*

3.2 Possibly tolerance and strongly tolerance solvability

The notions of the possibly tolerance and strongly tolerance solvability of interval max-min matrix equation (6) are defined in this section.

Definition 4.
(i) A matrix $B \in \boldsymbol{B}$ is a strongly tolerance matrix of interval max-min matrix equation (6) if there exists a matrix $X \in B(n,s)$ such that $A \otimes X \otimes C = B$ for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$.
(ii) Interval max-min matrix equation (6) is strongly tolerance solvable if there exists a matrix $B \in \boldsymbol{B}$ such that B is a strongly tolerance matrix of (6).

(iii) Interval max-min matrix equation (6) is possibly tolerance solvable if there exists a matrix $B \in \boldsymbol{B}$ such that for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ there exists $X \in B(n,s)$ such that $A \otimes X \otimes C = B$.

Lemma 6. A matrix $B \in \boldsymbol{B}$ is a strongly tolerance matrix of interval max-min matrix equation (6) if and only if it satisfies equality (7).

Proof. A matrix $B \in \boldsymbol{B}$ is a strongly tolerance matrix of (6) if and only if interval max-min matrix equation (6) with the constant right-hand side $\underline{B} = \overline{B} = B$ is strongly universally solvable, which is according to Theorem 3 equivalent to (7). □

We define the sequence of matrices $\{B^{(t)}\}_{t=0}^{\infty}$ as follows:

$$B^{(t)} = \begin{cases} \overline{B} & \text{for } t = 0, \\ \underline{A} \otimes X^*(\overline{A}, B^{(t-1)}, \overline{C}) \otimes \underline{C} & \text{for } t \geq 1. \end{cases} \tag{10}$$

Lemma 7. The following assertions hold true for the sequence $\{B^{(t)}\}_{t=0}^{\infty}$ defined by (10):
(i) The sequence $\{B^{(t)}\}_{t=0}^{\infty}$ is non-increasing.
(ii) There exists $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)}$.

Proof.
(i) By Theorem 1 (ii) and by the monotonicity of $\otimes$ we have

$$B^{(t+1)} = \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C} \leq \overline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \overline{C} \leq B^{(t)}.$$

(ii) It follows from (5) and (10), that $b_{ik}^{(t)} \in \{\overline{b}_{ik},\ i \in M,\ k \in R\} \cup \{\underline{a}_{ij},\ i \in M,\ j \in N\} \cup \{\underline{c}_{kl},\ k \in R,\ l \in S\}$ for each $i \in M, k \in R$ and for each $t \in \mathbb{N}_0$, i.e., at most $mn+mr+sr$ different values in each entry of $B^{(t)}$ can be considered. From part (i) it follows that the sequence $\{B^{(t)}\}_{t=1}^{\infty}$ is non-increasing, so the number of different matrices $B^{(t)}$ is bounded by $mr(mn+mr+sr)$. Consequently there exists $u \in \mathbb{N}_0$, $u \leq mr(mn+mr+sr)$ such that $B^{(u+1)} = B^{(u)}$. □

Theorem 8. *Let $B \in \boldsymbol{B}$ be a strongly tolerance matrix of interval max-min matrix equation* (6)*. Then the inequality $B \leq B^{(t)}$ is satisfied for each $t \in \mathbb{N}_0$.*

Proof. By mathematical induction on t
1. For $t = 0$ the inequality $B \leq \overline{B} = B^{(0)}$ trivially holds.
2. Suppose that $B \leq B^{(t)}$. We have

$$B = \underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C} \leq \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C} = B^{(t+1)}.$$

□

Theorem 9. *Interval max-min matrix equation* (6) *is strongly tolerance solvable if and only if there exists $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)} \in \boldsymbol{B}$.*

Proof. If $B^{(u+1)} = B^{(u)} \in \boldsymbol{B}$ for some $u \in \mathbb{N}_0$ then $\underline{A} \otimes X^*(\overline{A}, B^{(u)}, \overline{C}) \otimes \underline{C} = B^{(u)} \in \boldsymbol{B}$ which means that matrix $B^{(u)}$ is a strongly tolerance matrix of (6). Thus interval max-min matrix equation (6) is strongly tolerance solvable.

For the converse implication suppose that interval max-min matrix equation (6) is strongly tolerance solvable and $B \in \boldsymbol{B}$ is a strongly tolerance matrix of (6). According to Lemma 7 (ii) there exists $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)}$. Then

$$\underline{B} \leq B \leq B^{(u)} = B^{(u+1)} \leq \overline{B},$$

where the last inequality follows from Lemma 7 (i). Hence $B^{(u+1)} = B^{(u)} \in \boldsymbol{B}$. □

Theorem 9 enables us to give the following algorithm for checking the strongly tolerance solvability.

Algorithm ST
Input: $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$.

Output: 'yes' in variable st and strongly tolerance matrix B^* if the given interval max-min matrix equation is strongly tolerance solvable; 'no' in st otherwise.

begin

(i) $B^{(0)} = \overline{B},\ t = 0$;
(ii) $B^{(t+1)} = \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C}$;
(iii)**If** $\underline{B} \not\leq B^{(t+1)}$ **then** st :='no', go to **end**;
(iv)**If** $B^{(t+1)} = B^{(t)}$ **then** st :='yes', $B^* = B^{(t)}$, **go to end**;
(v) $t := t + 1$, **go to** (ii);

end

The following theorem deals with the complexity of checking the strong tolerance solvability. For the sake of simplicity, we suppose in the following theorem that $m = r = s = n$, i.e., all matrices are square of order n.

Theorem 10. *Algorithm ST decides whether the given interval max-min matrix equation of the form* (6) *is strongly tolerance solvable in* $O(n^7)$ *steps.*

Proof. The most time-consuming is step (ii) which requires $O(n^3)$ operations. The question which arises is the number of repetitions of the loop (ii)–(v) till the algorithm gives an answer. This number is bounded by the number of different matrices $B^{(t)}$, which is $O(n^4)$ (see the proof of Lemma 7 (ii)). Consequently, the complexity of Algorithm ST is $O(n^7)$.

□

Theorem 11. *Interval max-min matrix equation* (6) *is possibly tolerance solvable if and only if it is strongly tolerance solvable.*

Proof. Interval max-min matrix equation (6) is possibly tolerance solvable if and only if there exists a matrix $B \in \boldsymbol{B}$ such that interval max-min matrix equation with constant matrix $\underline{B} = \overline{B} = B$ is weakly universally solvable. According to Theorem 5, weak universal solvability is equivalent to universal solvability. Hence the weak possible tolerance solvability is equivalent to the existence $B \in \boldsymbol{B}$ such that the equality $\underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C}$ is satisfied. The assertion follows from Lemma 6. □

Example 2. Let $\mathcal{I} = [0, 10]$ and $(\mathcal{I}, \oplus, \otimes)$ be max-min algebra. Let us take

$$\boldsymbol{A} = \begin{pmatrix} [4,8] & [5,5] \\ [3,3] & [3,5] \\ [2,9] & [5,8] \end{pmatrix},\ \boldsymbol{B} = \begin{pmatrix} [3,9] & [2,10] & [1,4] \\ [2,5] & [3,4] & [2,5] \\ [2,9] & [1,3] & [3,4] \end{pmatrix},\ \boldsymbol{C} = \begin{pmatrix} [4,10] & [3,6] & [3,9] \\ [2,3] & [2,3] & [4,5] \end{pmatrix}.$$

We check whether the interval matrix equation $\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B}$ is strongly tolerance solvable.

Solution:
We compute the sequence $\{B^{(t)}\}_{t=0}^{\infty}$. We obtain

$$\{B^{(t)}\}_{t=0}^{\infty} = \left\{ \begin{pmatrix} 9 & 10 & 4 \\ 5 & 4 & 5 \\ 9 & 3 & 4 \end{pmatrix}, \begin{pmatrix} 3 & 3 & 4 \\ 3 & 3 & 3 \\ 3 & 3 & 4 \end{pmatrix}, \begin{pmatrix} 3 & 3 & 4 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}, \dots \right\}.$$

Since $B^{(4)} = B^{(3)} \geq \underline{B}$, according to Theorem 9 the given interval max-min matrix equation is strongly tolerance solvable. √

3.3 Conclusion

In this paper, we dealt with the solvability of interval matrix equations in max-min algebra. Max-min algebra is a useful tool for describing real situation in the economy and industry. In Example 1, the values a_{ij}, x_{jl}, and c_{lk} represent the capacities of corresponding connections. In economics, those values can represent for example the financial costs for the production or transporting of some products. In another example, a_{ij} represents a measure of the preference of the property P_i of some object before the property Q_j, similarly x_{jl} (c_{lk}) represent a measure of preference of property Q_j to property T_l (property T_l to property D_k). There are many possibilities for definition of more solvability concepts. Our main objective for the future will be studying another types of solvability of interval system with bounded and unbounded solution.

References

[1] Cuninghame-Green, R. A. (1979). *Minimax Algebra*. Lecture notes in Economics and Mathematical systems. Berlin: Springer.

[2] Draženská. E. & Myšková, H. (2017). Interval fuzzy matrix equations. *Kybernetika*, 53 (1), 99–112.

[3] Gavalec, M. & Plavka, J. (2010). Monotone interval eigenproblem in max-plus algebra. *Kybernetika*, 46 (3), 387–396.

[4] Myšková, H. (2005). Interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 403, 263–272.

[5] Myšková, H. (2006). Control solvability of interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 416, 215–223.

[6] Myšková, H. (2012). An iterative algorithm for testing solvability of max-min interval systems. *Kybernetika*, 48 (5), 879–889.

[7] Myšková, H. (2012). On an algorithm for testing T4 solvability of fuzzy interval systems. *Kybernetika*, 48 (5), 924–938.

[8] Myšková, H. (2012). Max-min interval systems of linear equations with bounded solution. *Kybernetika*, 48 (2), 299–308.

[9] Myšková, H. (2012). Robustness of interval Toeplitz matrices in fuzzy algebra. *Acta Electrotechnica et Informatica*, 12 (4), 56–60.

[10] Myšková, H. (2012). Interval eigenvectors of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica* 12 (3), 57–61.

[11] Myšková, H. (2012). Weak stability of interval orbits of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica* 12 (3), 51–56.

[12] Myšková, H. (2016). Interval max-plus matrix equations. *Lin. Algebra Appl.*, 492, 111–127.

[13] Myšková, H. & Plavka, J. (2019). On the solvability of interval max-min matrix equations. *Lin. Algebra Appl.*, Article in Press

[14] Olsder, G. J. et al. (1998). Course Notes: Max-algebra Aproach to Discrete Event Systems. *Algebres Max-Plus et Applications an Informatique et Automatique*, INRIA, 1998, 147–196.

[15] Plavka, J. (2012). On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices. *Discrete Appl. Math.*, 160, 640–647.

[16] Zimmermann, K. (1976). *Extremální algebra*. Praha: Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV.

The Problem of Back-test Overfitting in Quantitative Trading

Jiří Witzany[1]

Abstract. Quantitative investment strategies are often selected from a broad class of candidate models estimated and tested on historical data. Standard statistical techniques to prevent model overfitting such as out-sample back-testing turns out to be unreliable in the situation when selection is based on results of too many models tested on the holdout sample. There is an ongoing discussion how to estimate the probability of back-test overfitting and adjust the expected performance indicators like Sharpe ratio in order to reflect properly the effect of multiple testing. We propose a consistent Bayesian approach that consistently yields the desired robust estimates based on an MCMC simulation. The approach is tested on a class of technical trading strategies where a seemingly profitable strategy can be selected in the naïve approach.

Keywords: Back-test, multiple testing, bootstrapping, cross-validation, investment strategy, Sharpe ratio, Bayesian probability, MCMC

JEL Classification: G1, G2, C5, G24, C11, C12, C52
AMS Classification: 91G70, 91G60, 91B60

1 Introduction

The problem of back-test overfitting appears mainly in two important econometric research areas: testing and selection of factors explaining asset returns (see e.g. Harvey and Liu [3], [4] or [6]) and selection of investment strategies (see e.g. White [12] or Bailey et al. [2]). Our focus is the investment strategy selection problem arising when many strategies are developed and tested on historical data in order to find a "performing" one. The selection process can be realized by an individual researcher or institution, or latently by a set of researchers investigating various strategies and publishing only the promising ones. The latter approach is more common for theoretical research while the former, easier to control, would be typical for a quantitative investment firm.

We are going to formalize and investigate the problem of strategy selection based on a large set of candidates By a strategy, we mean a function prescribing portfolio positions (long or short) given all available information at time t and preserving the standard self-financing condition, i.e. the same portfolio value before and after the asset reallocation.. Consider strategies $\mathcal{S}_1, \dots, \mathcal{S}_K$ that are backtested and evaluated over a historical period with T_1 (e.g. daily) returns $r_{k,t}, k = 1, \dots, K, t = 1, \dots, T_1$. Note that the strategies could have been developed on another preceding training period and backtested or validated on the $\{1, \dots, T_1\}$ period. Another possibility that we use in the empirical study is that one considers a number of expertly proposed, e.g. technical, strategies that are just evaluated on the backtest period. Based on the historical data we estimate the (annualized) sample means m_k, standard deviations s_k, or Sharpe ratios SR_k, and given a criterion select the "best" strategy $\mathcal{S}_b$. Of course, the key question is what can be realistically expected from the best strategy if implemented in the future period of length T_2 (see Figure 1).

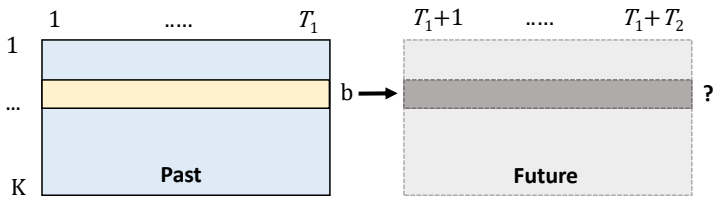


Figure 1. Future performance of the best strategy selected based on the past data

It is obvious that is not correct to use the historical performance of the selected strategy such as the Sharpe ratio as an unbiased predictor of the future performance. It also not correct to interpret the classical p-value of as

[1] University of Economics, Faculty of Finance and Accounting, Department of Banking and Insurance, W. Churchill Sq. 4, 130 67, Prague, Czech Republic, +420 224 095 174, e-mail: jiri.witzany@vse.cz

the probability that the selected strategy is not a truly profitable strategy. Our goal is to study methods how to adjust the classical indicators or estimate the probability of backtest overfitting (Bailey et al. [2]).

We are going propose a Bayesian methodology that allows us to simulate many times the in-sample (IS) selection and the out-of-sample (OOS) realization process (Figure 1) in order to address the questions formulated above. We consider the method consistent since it repeats, in a Monte Carlo simulation, the process of strategy selection and ex post performance realization based on Bayesian parameters extracted from given data, of course conditional on a data generating model. We will provide an overview of several methods proposed in literature that will be compared with our proposed strategy in an empirical study.

2 An Overview of the Existing Approaches

There are several relatively simple classical methods how to adjust p-values in order to accommodate the multiple test. More advanced and computationally demanding methods are based on various approaches to bootstrapping and simulation of the past and future returns.

Classical approaches

To test significance of a single strategy, for example $\mathcal{S}_b$ with sample mean return m_b and standard deviation s_b observed over T periods, the classical approach is to calculate the t-ratio $TR = \frac{m_b}{s_b/\sqrt{T}}$ and the two-sided p-value

$$p^S = \Pr[|X| > TR] \tag{1}$$

where X is a random variable following the t-distribution with $T-1$ degrees of freedom. The implicit assumption is that the returns are i.i.d. normal. If the p-value p^S happens to be small enough, e.g. below 5% or 1%, then one tends to jump to the conclusion that a strategy with significantly positive returns has been discovered.

The problem of the process of selecting the best strategy, or alternatively testing a number of strategies until we find a significant one, is that the correct p-value should be (Harvey, Liu [5]) reflecting the fact that we are selecting the strategy with the best t-ratio TR out of K strategies $p^M = \Pr[\max\{|X_k|, k = 1, \dots, K\} > TR]$. Here X_k are k random variables following the t-distribution with $T-1$ degrees of freedom (corresponding to the sample t-ratios of the of K strategies). It can be noted (Harvey, Liu [5]) that if the variables were independent then we could find a simple relationship (also called Šidák's adjustment) between the single and multiple test p-values:

$$p^M = 1 - \prod_k \Pr[|X_k| \le TR] = 1 - (1-p^S)^K = Kp^S - \binom{K}{2}(p^S)^2 + \cdots$$

(Harvey, Liu [5]) provide an overview of simple adjustment methods, such as Bonferroni's adjustment $p^M = \min\{Kp^S, 1\}$, Holm's or Benjamini, Hochberg, Zekutieli (BHY) adjustments based on the ordered sequence of the single test p-values $p_1^S, \dots, p_K^S$ for all the strategies. The weak point of all those methods is the assumption of independence since the tested strategies are often closely related (e.g. of the same type with varying parameters).

We are also proposing and will test a numerically relatively simple and efficient method based on an estimation of the covariance matrix Ω of the returns and numerically generating the distribution of $\max\{|X_k|, k = 1, \dots, K\}$ conditional on the null hypothesis $m_k = 0$ for all k where X_k are K random variables following the t-distribution with $T-1$ degrees of freedom (or alternatively standard normal for a large T) and with covariances given by Ω.

Note that the classical Sharpe ratio can be easily calculated given the t-ratio and vice-versa $SR_k = \frac{m_k}{s_k} = \frac{TR_k}{\sqrt{T}}$. The Sharpe ratio is usually annualized as follows $SR_k^a = \frac{m_k}{s_k}\sqrt{T_a} = TR_k\sqrt{\frac{T_a}{T}}$, where T_a is the number of observation periods in a year, e.g. 252 in case of daily returns. According to (1) the maximal acceptable p-value level can be easily translated to a minimum required Sharpe ratio.

Generally, given a selected strategy with in-sample (based on the backtest data) Sharpe ratio SR_{IS} the question is what is the expected (out-of-sample) Sharpe ratio $E_0[SR_{OOS}]$ over a future, e.g. one-year, period. Here, $E_0[.]$ denotes the expectation given all the information available today, in particular given the in-sample performance like SR_{IS}, the number of strategies from which the best one was selected, the relationship between the strategies, the underlying asset return process properties, etc. The Sharpe ratio haircut is then defined as the percentage we need to deduct from the in-sample Sharpe ratio to get a realistic estimate of the future performance,

$$HC = 1 - \frac{E_0[SR_{OOS}]}{SR_{IS}}. \tag{2}$$

Harvey and Liu [5] note that the rule-of-thumb haircut used by the investment industry is 50% but that, according to their analysis it significantly depends on the level of the in-sample Sharpe ratio and the number of strategies. They propose to use the relationship between the single and multiple test p-values in order to estimate the haircut Sharpe ratio. Their estimate of the annualized expected Sharpe ratio ESR_{HL} is based on the idea that its corresponding single test p-value should be equal to the adjusted multiple test p-value p^M, i.e. $p^M = \Pr\left[|X| > ESR_{HL}\sqrt{\frac{T}{T_a}}\right]$, $SR_{HL} = F^{-1}(p^M/2)\sqrt{\frac{T_a}{T}}$, where X is a random variable following the t-distribution with $T-1$ degrees of freedom and F is its cumulative distribution function. The haircut is then calculated according to (2). The haircut estimation, of course depends, on the p-value adjustment method as Bonferroni, Holm's, BHY, or the general one we have suggested above. Although the estimation is obviously directionally correct, it is not obvious why this approach should yield a consistent estimate of the expected Shar ratio $E_0[SR_{OOS}]$ and of the corresponding haircut. We are going to compare the different haircut estimates in the simulation study outlined below.

Stationary Bootstrap

In order to simulate the past and the future returns we are going to consider a bootstrapping and a cross-validation approach. The *stationary bootstrap* proposed and analyzed in White [12], Sullivan et al. [11] and Politis, Romano [9] is applied to the underlying asset returns assumed to be strictly stationary and weakly dependent time-series to generate a pseudo time series that is again stationary.

Formally, we generate new sequences of the underlying asset returns $\{u_{\Theta(i)}; i = 1, .., \tilde{T}\}$ where $u_1, \dots, u_{T_1}$ is the original series of observed returns and $\Theta(i) \in \{1, \dots, T_1\}$. In order to implement the bootstrap we need to select a smoothing parameter $0 < q = 1/b < 1$, where b corresponds to the mean length of the bootstrapped blocks, for example $q = 0.1$ proposed by Sullivan et al. [11]. A bootstrapped sequence is obtained by drawing randomly $\Theta(1) \in \{1, \dots, T\}$, and for $i = 2, \dots, \tilde{T}$ setting $\Theta(i) = \Theta(i-1) + 1$ with probability $1 - q$ or randomly drawing a new block starting position $\Theta(i) \in \{1, \dots, T_1\}$ with probability q. If it happens that $\Theta(i) > T_1$ then we draw random $\Theta(i) \in \{1, \dots, T_1\}$.

Next, given a bootstrapped sequence of the underlying asset returns we need to apply strategies $\mathcal{S}_1, \dots, \mathcal{S}_K$ to get the strategies' bootstrapped returns $\tilde{r}_{k,t}, k = 1, \dots, K, t = 1, \dots, T_1$. Note that since the strategies' decisions are often built based on the past we generally need to have a longer series of the bootstrapped asset returns, $\tilde{T} > T_1$ where T_1 is the length of the in-sample period. Then we evaluate our desired performance indicator values (mean return, Sharpe ratio, etc.) $\tilde{f}_k$. Let f_k^* denote the performance indicators of the original series of returns. According to White [12], under certain mild theoretical assumptions, the empirical distribution of the B bootstrapped values $\tilde{V}_j = \max_{k=1,\dots,K}(\tilde{f}_k - f_k^*)$ for $j = 1, \dots, B$ asymptotically converges to the distribution of the best strategy performance indicator under the null hypothesis H_0 that all the strategies have zero performance. I.e., obtaining B bootstrapped values $\{\tilde{V}_j; j = 1, \dots, B\}$ we can test H_0 by calculating the empirical p-value $\Pr\left[\left|\tilde{V}_j\right| > f_b^*\right]$, where b is the index of the best strategy and $f_b^* = \max_k f_k^*$.

The bootstrap technique can be also used to analyze the relationship between IS and OOS Sharpe ratio (or another indicator) generating series of strategy returns over a time period $1, \dots, T_1$ selecting the best strategy $\mathcal{S}_b$ with in-sample performance SR_{IS} and then looking on its out-of-sample performance SR_{OOS} over the following period $T_1 + 1, \dots, T_1 + T_2$. Note that in this case the original bootstrapping has to be done over a period of length $\tilde{T} > T_1 + T_2$. Then we can compare the mean SR_{OOS} against the mean SR_{IS}, or conditional on certain level of SR_{IS}. We may also bootstrap the OOS returns for the actually selected strategy $\mathcal{S}_b$ (based on the real dataset). However, particularly in this case, it is obvious that even a truly positive strategy that is using medium-term or long-term trends to make good predictions does not have to work on the bootstrapped series of returns where the future and past returns of the original series are to large extent mixed up. Therefore, the estimated conditional SR_{OOS} may easily lead to a false rejection of a good strategy.

Combinatorial Symmetric Cross-Validation

Another disadvantage of the stationary bootstrap technique is that it cannot be applied if we are given only the strategy returns but not details on the strategies themselves. The stationary bootstrap is also problematic if the strategies are not technical ones and use a number of additional, possibly lagged, explanatory factors. This is not the case of the *combinatorial symmetric cross-validation* (CSCV) (Bailey et al. [1], [2]) utilizing only the matrix of the strategies' returns $M = \{r_{k,t}, k = 1, \dots, K, t = 1, \dots, T_1\}$. The idea is to split the time window of length $T_1 = SN$ into S blocks of length N where S is even and draw combinations of $S/2$ blocks. The submatrix J formed by joining $T_1/2$ rows of M corresponding to the selected time indices in the original order then represents an in-sample dataset of returns where the best performing strategy can be selected while the complementary $K \times T_1/2$

submatrix $\bar{J}$ represents the out-of-sample returns. The sampling can be done with or without replacement. Since there are $\binom{S}{S/2}$ combinations we can form sufficiently many different combinations with replacement as long as S is sufficiently large, e.g. at least 16.

Bailey et al. [1], [2] propose to use the technique to estimate specifically the Probability of Backtest Overfitting (PBO) defined as the probability that the best IS selected strategy performs below the average OOS. More precisely, for K strategies $\mathcal{S}_1, \dots, \mathcal{S}_K$, $PBO = \Pr[Rank_{OOS}(X) < K/2 | Rank_{IS}(X) = 1]$. The PBO indicator as well as the Sharpe ratio haircut can be estimated using sufficiently many cross-validation pairs of the IS/OOS datasets $\langle J, \bar{J} \rangle$. However, it is obvious that the estimates are biased introducing a negative drift into the OOS order of the strategies. For example, if all the strategies represented just pure noise with mean returns over the full time interval $\{1, \dots, T_1\}$ close to zero, then for an IS/OOS combination $\langle J, \bar{J} \rangle$ the best strategy IS return $\bar{r}_{b,J}$ implies that the complementary OOS return $\bar{r}_{b,\bar{J}} \approx -\bar{r}_{b,J}$ would be probably the worst on $\bar{J}$. We will demonstrate the effect in the empirical part. The cross-validation technique also cannot be used, due to this property, to estimate the OOS Sharpe ratio or mean for a particular selected strategy. We can just estimate the overall PBO or Sharpe ratio haircut keeping in mind that the estimations incorporate a conservative bias. The cross-validation as well as the bootstrapping approach cannot be easily used to estimate the False Discovery Rate (or equivalently the Family Wise Error Rate if the best IS strategy is automatically declared to be a discovery) since it is not clear how to identify true and false discoveries given a CSCV simulation. This could be possibly done by testing significance of OOS performance involving an ad hoc probability level. We are going to show that all the indicators of interest can be consistently estimated in a Bayesian set up we are going to outline below.

3 Bayesian Simulation Approach

The Bayesian approach will be based on the scheme following Figure 1. First, of course we need to specify a model defining the return generating process with unknown parameters Θ for the observed strategy returns $\{r_{k,t}, k = 1, \dots, K, t = 1, \dots, T\}$ where $T = T_1$ or $T = T_1 + T_2$. Then the plan is to use a Bayesian technique, in particular the Markov Chain Monte Carlo (MCMC) simulation in order to extract the posterior distribution of the model parameters Θ. Finally, simulate the matrices of IS and OOS returns over desired time intervals $1, \dots, T_1$ and $T_1 + 1, \dots, T_1 + T_2$. The Monte Carlo (MC) is done in two steps always selecting the parameters Θ from the posterior distribution and then generating K series of $T_1 + T_2$ returns according to the model. The simulated IS returns can be used to select the best strategy and the OOS returns to measure its future performance. The average haircut or average relative rank can be easily estimated as in case of the stationary bootstrap (see **Chyba! Nenalezen zdroj odkazů.**).

We are going to consider two models, the simple one assumes that the returns are multivariate normal with unknown covariance matrix and means while the second incorporates unknown indicators of truly profitable strategies allowing us to estimate consistently the false discovery rate (FDR) etc. The second model follows an idea of Scott and Berger [10], also mentioned in Harvey et al. [7], nevertheless, in both cases the model is formulated only for observed mean returns and without considering a correlation structure of returns. It should be emphasized that our focus is to analyze the impact of backtest overfitting assuming that the strategies' cross-sectional returns behave in a relatively simple and stable way over time similarly to the classical, bootstrapping or cross-validation approaches. One could certainly come up with state-of-the art models incorporating jumps, switching regimes, stochastic variances or even dynamic correlations. These improvements would make the methodology computationally difficult to manage with results probably even more conservative compared to the approaches we are going to consider below.

The Naïve Model 1

To set up the naïve model we assume that the cross sectional strategy returns are multivariate normal $\boldsymbol{r}_t = \langle r_{1,t}, \dots, r_{K,t} \rangle \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and that the observations over time are independent. Given $\text{data} = \langle \boldsymbol{r}_t \rangle$, i.e. the matrix of back test returns, and possibly some priors for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we can find the posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \text{data})$ using the standard Gibbs MCMC sampler.

Specifically, the iterative sampling is given by $p(\boldsymbol{\mu} | \boldsymbol{\Sigma}, \text{data}) = \varphi\left(\boldsymbol{\mu}; \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t, \frac{1}{T}\boldsymbol{\Sigma}\right)$ and $p(\boldsymbol{\Sigma} | \boldsymbol{\mu}, \text{data}) = IW(\boldsymbol{\Sigma}; T, S)$, where $S = \sum_{t=1}^{T} (\boldsymbol{r}_t - \boldsymbol{\mu})' (\boldsymbol{r}_t - \boldsymbol{\mu})$ is the scale matrix (i.e. the covariance matrix times T) and IW is the Inverse Wishart distribution. For example, Matlab allows to sample from the distributions and the posterior distribution may be obtained quite efficiently (e.g. 10 000 runs of the sampler). The sampler assumes the non-informative prior on the means, $p(\boldsymbol{\mu}) \propto 1$, and the standard improper prior on the covariance matrix $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{K+1}{2}}$.

Given the extracted posterior distribution $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\text{data})$ the parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ can be now easily sampled in order to get the empirical distribution of the selected strategy performance. However, in the process of selecting the best strategy we do not know $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ but only a time series of the back tested returns with cross sections from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$**.** Based on the time-series the "best" strategy $\mathcal{S}_b$ is selected. Our key question is about its expected forward-looking performance, e.g. μ_b or SR_b. Therefore, we need to run the following Monte Carlo simulation in order to sample faithfully the empirical distribution of the performance indicators:

1. Sample $\langle \boldsymbol{\mu}, \boldsymbol{\Sigma} \rangle$ from $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\text{data})$.
2. Sample independently $T_1 + T_2$ cross sections $\boldsymbol{R_t} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
3. Determine the index of the best strategy b based on the back-test statistics calculated from the matrix of back-tested returns $\boldsymbol{R} = \langle \boldsymbol{R_t} \rangle$ for $t = 1, .., T_1$.
4. Calculate and store the performance indicators, $\hat{\mu}_b$, $\widehat{SR}_b$, on the OOS period $T_1 + 1, \dots, T_2$. Alternatively, store the selected strategy "true" performance indicators, i.e. μ_b, SR_b.

The MCMC estimation of the multivariate normal distribution parameters is known to converge relatively fast (see e.g. Lynch [8]). But we need to apply a burn-out period based on a simple diagnostics, e.g. based on average of the vector of mean returns $\boldsymbol{\mu}$. The simulated posterior distribution of the desired performance indicators (after removing the burn-out period) then tells us what are the mean, median, confidence intervals, or Bayesian probabilities that the true performance is positive, or above any given minimum threshold. The ratio between the ex-post and ex-ante performance indicators also give us an estimate of the "backtest overfitting haircut."

Model 2 – Bimodal Means Distribution

In order to capture the situation when most strategies are random and only some positive (non-zero) assume that there are in addition latent indicators $\gamma_i \in \{0,1\}$ so that the mean of strategy i is $\mu_i^* = \gamma_i \mu_i$. Therefore, the row vector of returns has the distribution $\boldsymbol{r}_t = [r_{1,t}, \dots, r_{K,t}] \sim N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$**.** Here we need to assume a prior Bernoulli distribution for $\gamma_i \sim Bern(1 - p_0)$ and $\mu_i \sim N(m_0, V_0)$. It means that the Bayesian distribution of the means is bimodal with a large probability mass on 0 and the other mode being normal with prior mean $m_0 > 0$ and variance V_0. The Gibb's sampler can be modified as follows:

1. Given $\boldsymbol{\mu}, \boldsymbol{\gamma}$, set $\mu_i^* = \gamma_i \mu_i$, and estimate $\boldsymbol{\Sigma}$ as above, i.e. $p(\boldsymbol{\Sigma}|\boldsymbol{\mu}, \boldsymbol{\gamma}, \text{data}) = IW(\boldsymbol{\Sigma}; T, S)$, where $S = \sum_{t=1}^{T} (\boldsymbol{r}_t - \boldsymbol{\mu}^*)' (\boldsymbol{r}_t - \boldsymbol{\mu}^*)$.
2. Given $\boldsymbol{\Sigma}, \boldsymbol{\gamma}$, estimate $\boldsymbol{\mu}$. Set $A = \frac{1}{T}\boldsymbol{\Sigma}$, $\Gamma = diag(\boldsymbol{\gamma})$, $\boldsymbol{m} = \frac{1}{T}\sum_{t=1}^{t} \boldsymbol{r}_t$, $\boldsymbol{m}_0 = [m_0, \dots, m_0]$, $D = diag([V_0, \dots, V_0])$, where $diag$ creates a matrix with diagonal elements given by the vector in the argument, and sample $p(\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{\gamma}, \text{data}) = \varphi(\boldsymbol{\mu}; (\Gamma A^{-1}\Gamma + \mathrm{D}^{-1})(\Gamma \mathrm{A}^{-1} m + D^{-1}\boldsymbol{m}_0), (\Gamma A^{-1}\Gamma + \mathrm{D}^{-1})^{-1})$.
3. Given $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$, estimate $\boldsymbol{\gamma}$. For $i = 1, \dots, K$ set Γ_0 equal to Γ with the exception of the diagonal element $\Gamma_0(i,i) = 0$, and Γ_1 setting $\Gamma_1(i,i) = 1$. Let $L_0 = \exp\left(\frac{-1}{2}\left((\Gamma_0 \boldsymbol{\mu} - \boldsymbol{m})' A^{-1} (\Gamma_0 \boldsymbol{\mu} - \boldsymbol{m})\right)\right)(1 - p_0)$, $L_1 = \exp\left(\frac{-1}{2}\left((\Gamma_1 \boldsymbol{\mu} - \boldsymbol{m})' A^{-1} (\Gamma_1 \boldsymbol{\mu} - \boldsymbol{m})\right)\right) p_0$, $\tilde{p} = \frac{L_1}{L_0 + L_1}$, and finally sample $\gamma_i \sim Bern(\tilde{p})$.

4 A Numerical Study and Conclusions

Following Sullivan et al. [11] and other studies we have used 1000 daily S&P 500 values and returns for the period 5.6.2009 – 24.5.2013. The period has been selected with the purpose to find at least one strategy with a higher mean return. As in Sullivan et al. [11] we have applied the filter, moving average, support and resistance rules with varying parameters. We have selected randomly 200 strategies with the condition that the daily profit series are not collinear and applied the backtest adjustment methods described above with results in Table 1.

	Adjusted p-value (FDR)	**Ex ante av. SR/mean**	**Adjusted expected SR/mean**	**SR/mean hair cut**	**Probability of loss**	**Mean OOS rank**	**PBO**
Boferroni m.	1.00	1.199	0	100%	-	-	
Šidák's corr.	0.968	1.199	0.02	98.3%	-	-	
M.norm. MC	0.352	1.199	0.4668	61%	-	-	
Stat. boots.	0.728	1.110 / 0.194	0.297 / 0.051	73.2% / 74%	0.328	55%	0.444
CSCV	-	1.382 / 0.244	0.336 / 0.058	75.7% / 76.4%	0.371	66.8%	0.323
Bayes mod. 1	-	2.102 / 0.371	1.014 / 0.180	51.8% / 51.5%	0.171	75.7%	0.168
Bayes mod. 2	0.549	1.201 / 0.213	0.211 / 0.037	82.5% / 82.5%	0.380	60%	0.395

Table 1. Summary of the backtest overfitting tests

We have also modified the empirical test enlarging artificially the profitability of one of the strategies or de-meaning the strategies. Our testing have shown that the classical methods to adjust single test p-values for the effect of multiple testing when selecting a trading strategy out of many possibilities like Bonferroni, Holms, or BHY work relatively well but provide very conservative estimations due to their approximate nature. Certain improvement can be achieved applying the independence based multiple test p-value (Šidák's) adjustment or the proposed multivariate normal MC simulation method. The derived expected SR and the related haircut proposed by Harvey, Liu [5] are rather heuristic and in our view not well theoretically founded. The stationary bootstrap method proposed by Sullivan et al. [11] provides a consistent p-value adjustment. However, if used in a two-stage simulation it may damage functionality of a positive strategy depending on medium/long term trends due to the mixing bootstrap algorithm. It also turns out to be computational the most demanding since all strategies must be replicated for each sequence of bootstrapped asset prices. Moreover, it cannot be used if the strategies are not known or depend on other economic series. The CSCV method (Bailey et al. [2]) is relatively computational efficient and provides good results if the mean returns of the strategies are well diversified. However, if the strategies' mean returns are all close to zero then the method gives negatively biased results. On the other hand, it appears overoptimistic if one strategy stands high above the others. Finally, we have proposed and investigated two Bayesian methods, the naïve one based on the simple assumption that the returns are multivariate normal, and the second extended with latent variables indicating zero and nonzero mean return strategies. While the naïve model gives mixed results, the second provides, according to our empirical study, the most consistent results and a useful tool to analyze properly the issue of backtest overfitting. Besides the probability of loss and backtest overfitting (PBO) it estimates the posterior probabilities whether each individual strategy is a true discovery and at the same time the probability making a true discovery (and the complementary FDR) selecting the best one.

Acknowledgements

The author acknowledges the financial support from the Czech Science Foundation Grant Number 18-05244S (Innovation Approaches to Credit Risk Management).

References

[1] Bailey, D. H., Borwein, J. M., de Prado, M. L., & Zhu, Q. J. (2014). Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance. Notices of the AMS, 61(5), 458-471.

[2] Bailey, D. H., Borwein, J. M., Lopez de Prado, M., & Zhu, Q. J. (2016). The Probability of Backtest Overfitting. Journal of Computational Finance, Forthcoming. Available at SSRN: https://ssrn.com/abstract=2840838.

[3] Harvey, C. R. and Liu, Y., Multiple Testing in Economics (2013). Available at SSRN: https://ssrn.com/abstract=2358214

[4] Harvey, C. R., & Liu, Y. (2014). Evaluating trading strategies. The Journal of Portfolio Management, 40(5), 108-118.

[5] Harvey, C. R., & Liu, Y. (2015). Backtesting. The Journal of Portfolio Management, 42(1), 13-28.

[6] Harvey, C. R. and Liu, Y. (2017). Lucky Factors. Available at SSRN: https://ssrn.com/abstract=2528780 or http://dx.doi.org/10.2139/ssrn.2528780 , June 1

[7] Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the Cross-Section of Expected Returns. The Review of Financial Studies, 29(1), 5-68.

[8] Lynch, S. M. (2007). Introduction to Applied Bayesian Statistics and Estimation for Social Scientists. Springer, pp. 359.

[9] Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. Journal of the American Statistical association, 89(428), 1303-1313.

[10] Scott, J. G., & Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. Journal of statistical planning and inference, 136(7), 2144-2162.

[11] Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. The journal of Finance, 54(5), 1647-1691.

[12] White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126

Voigt Distribution and Its Heavy-tail Modeling Ability for Cryptocurrencies

Jaromir Kukal[1], Quang Van Tran[1]

Abstract. Voigt distribution is a convolution of a Cauchy distribution and a Gaussian distribution. This distribution has been so far widely used in technical disciplines. The inclusion of the Cauchy distribution into the mixture makes it a possible alternative for modeling heavy tail properties commonly present in financial data. In our contribution, the close form of distributional characteristics of the distribution like characteristic function (CF), probability density function (PDF) and cumulative distribution function (CDF) is derived. Due to their numerical complexity, these functions do not provide any practical usage for the maximum likelihood estimation (MLE) of its parameters. Therefore, we propose the use of Fast Fourier Transform (FFT) to make the parameter estimation via MLE technique more numerically feasible. This approach is then verified on the returns of cryptocurrencies which tend to have more heavy tails than those of ordinary heavy tail distributions. The results we obtained with this method are quite interesting.

Keywords: Voigt distribution, Fast Fourier Transform, MLE parameter estimation, Cryptocurrencies.

JEL Classification: C58, C65
AMS Classification: 62F10, 65T50

1 Introduction

Heavy tail property of financial asset returns is well-known and many alternatives have been proposed to capture it (Nolan [7], Eberlein and Keller [5], Theodossiou [8]). So far the results have been mixed therefore there is a room for improvement. To bring our effort to the search for an answer to this question, we examine the ability of Voigt distribution for this task. Voigt distribution is a convolution of a Cauchy distribution and a Gaussian distribution. The presence of the Cauchy distribution in the mix may introduce a new feature absent in a Gaussian distribution to model heavy tail property in financial data. Though the close form of probability density function exists, the numerical complexity of its evaluation makes the parameter maximum likelihood estimation procedure using directly the PDF almost impossible. Therefore, in this paper, we propose the use of Fast Fourier Transform for MLE approach as a more accessible alternative. This approach will be tested on the modeling daily returns of four most liquid cryptocurrencies: Bitcoin, Litecoin, Ripple (XRP), and Ethereum. As the returns of cryptocurrencies are highlzy volatile, they might be the result of the two random processes combined. The results we obtained with this method will be compared with those of each component of the convolution as well as their general version Student t-distribution to show the quality of our approach as well as the suitability of Voigt distribution for modeling extremely heavy tail financial data.

2 Voigt Distribution Primer

The Voigt distribution is a convolution of Gaussian and Cauchy distributions which is widely used in many technical disciplines (see Armstrong [1]). Let's have two random variables: one is normally distributed $A \sim \mathrm{N}(\mu, \sigma^2)$ and the other follows Cauchy distribution $B \sim \mathrm{C}(0, \gamma)$, then the random variable as the sum of these two variables is said to have the Voigt distribution $X = A + B \sim \mathrm{V}(\mu, \sigma, \gamma)$ with three parameters μ, σ, γ. By definition the Voigt distribution is a convolution

$$\mathrm{V}(\mu, \sigma, \gamma) = \mathrm{N}(\mu, \sigma^2) * \mathrm{C}(0, \gamma) \tag{1}$$

and its probability density function can be expressed as

$$\mathrm{f}(x, \mu, \sigma, \gamma) = \int_{-\infty}^{+\infty} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(\xi-\mu)^2}{2\sigma^2}\right) \cdot \frac{\gamma \mathrm{d}\xi}{\pi(\gamma^2 + (x-\xi)^2)}. \tag{2}$$

[1] Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2, 120 00 - Czech Republic, jaromir.kukal@fjfi.cvut.cz, tranvqua@fjfi.cvut.cz

Formula (2) does not provide us too much direct and effective usage. A more straightforward way to make use of properties of the Voigt distribution comes from its characteristic function[1].

2.1 Characteristic Function

Using parameters $\sigma, \gamma > 0$, $\mu, t \in \mathbb{R}$, and the well known convolution theorem, we directly obtain the characteristic function of the Voigt distribution as

$$\psi(t) = \mathrm{E}\exp(\mathrm{j}tX) = \exp(-\mathrm{j}\mu t - \sigma^2 t^2/2 - \gamma|t|). \tag{3}$$

Without location parameter μ, this CF is an even real function of t, but non–differentiable at $t = 0$. Generally, the Voigt distribution is symmetric around its median μ. However, the expected value and variance do not exist. It has two trivial limit cases

$$\psi_{\mathrm{Gauss}}(t) = \lim_{\gamma\to 0_+} \psi(t) = \exp(-\mathrm{j}\mu t - \sigma^2 t^2/2), \tag{4}$$

$$\psi_{\mathrm{Cauchy}}(t) = \lim_{\sigma\to 0_+} \psi(t) = \exp(-\mathrm{j}\mu t - \gamma|t|) \tag{5}$$

as CFs of the Gaussian and Cauchy distributions respectively shifted by μ.

2.2 Explicit Determination of PDF and CDF

The explicit enumeration of Voigt PDF is possible using a confluent hypergeometric function ${}_1\mathrm{F}_1$, which is defined in the whole complex domain $\mathbb{C}$ with similar properties as complex exponential. Let's define

$$z = \frac{x - \mu + \mathrm{j}\gamma}{2^{1/2}\sigma}. \tag{6}$$

The complex extension of the error function is

$$\mathrm{erf}(z) = 2z\pi^{-1/2}{}_1\mathrm{F}_1(1/2; 3/2; -z^2), \tag{7}$$

and the Faddeeva function is

$$\mathrm{w}(z) = \exp(-z^2)(1 - \mathrm{erf}(-\mathrm{j}z)). \tag{8}$$

The explicit form of Voigt distribution PDF is

$$\mathrm{f}(x, \mu, \sigma, \gamma) = \frac{\Re(\mathrm{w}(z))}{(2\pi)^{1/2}\sigma}. \tag{9}$$

The corresponding CDF of Voigt distribution can be expressed using a generalized hypergeometric function ${}_2\mathrm{F}_2$ in the complex domain $\mathbb{C}$ as

$$\mathrm{F}(x, \mu, \sigma, \gamma) = \Re\left(\frac{1 + \mathrm{erf}(z)}{2} + \frac{\mathrm{j}z^2}{\pi}{}_2\mathrm{F}_2(1, 1; 3/2, 2; -z^2)\right). \tag{10}$$

The main advantage of the generalized hypergeometric functions mentioned above is their implementability in MATLAB for complex argument with function `hypergeom(a,b,x)`. Unfortunately, the direct evaluations of Voigt PDF (9) and CDF (10) are very slow and realizable only for $|z| \leq 5$ due to increasing rounding errors.

3 Numeric Aspects of PDF and CDF Evaluation

The evaluation of Voigt distribution PDF and CDF according to (9,10) is very slow and numerically unstable. The PDF of the Voigt distribution can be obtained from its characteristic function (3) using inversion formula

$$\mathrm{f}(x, \mu, \sigma, \gamma) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-\mathrm{j}\mu t - \sigma^2 t^2/2 - \gamma|t| - \mathrm{j}tx)\mathrm{d}t \tag{11}$$

and the Fast Fourier Transform (FFT) (see Brigham [3]). We suggest to apply N–point symmetric scheme with step $\Delta t > 0$ for $N = 2^{24}, \Delta t = 3 \times 10^{-4}$. The numeric procedure of PDF evaluation begins with the enumeration of $\psi(t_k)$ at $t_k = (k - N/2)\Delta t$, for $k = 0, \ldots, N - 1$. Then series $\psi(t_k)$ is converted to series f_k using FFT and

[1] The shift parameter of Voigt distribution can be assigned to any one of two component distributions without the loss of generality as it is the sum of individual shifts by definition of a characteristic function of sum of these two random variables, see equation (3).

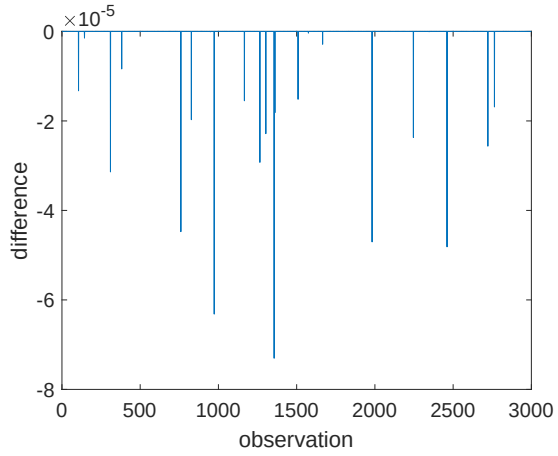


Figure 1 The differences in values of PDF computed by analytic and FFT based approaches

the values of f_k at x_k form a look–up table to obtain the values of PDF for given x via linear interpolation. The corresponding CDF of Voigt distribution can be easily obtained by integration

$$\mathrm{F}(x,\mu,\sigma,\gamma) = \frac{1}{2} + \int_{-\infty}^{x} \mathrm{f}(\xi,\mu,\sigma,\gamma)\mathrm{d}\xi. \tag{12}$$

Numerically, applying the trapezoidal rule to the PDF look–up table, we obtain values of F_k at x_k which can be used as another look–up table for the value of CDF at a given x also through linear interpolation. The applicability of this approach is examined as follows. A hundred samples of length 3000 with $\mu = 10^{-2}, \sigma = 1.3, \gamma = 0.6$. The average time for evaluation of Voigt PDF with FFT is $1.4632s$ (our PC parameters: CPU i5, RAM 4GB, 64 bit OS). The average time for computation Voigt PDF analytically with the same PC is $16.2961s$. Regarding the precision of the method, the difference of each evaluation can be observed in Figure 1. The total deviation in the values of PDF calculated by these two methods is roughly a half of a per mille (the exact value is 5.1547×10^{-4}).

4 Student t-Distribution and Cauchy Distribution

In this part, we briefly introduce Student t-distribution and Cauchy distribution. The Student t-distribution is the general case of both Cauchy distribution and Gaussian distribution. It has three parameters: location μ, scale σ and the so called the number of degrees of freedom ν. The third parameter ν defines the tail property of this distribution. The higher value ν attains, the lower value the kurtosis of this distribution is. When $\nu \to \infty$, t-distribution becomes Gaussian distribution. Its PDF is

$$\mathrm{f}(x,\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \tag{13}$$

and the corresponding CDF is

$$\mathrm{F}(x,\mu,\sigma,\nu) = \frac{1}{2} + \frac{x-\mu}{\sigma}\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2\mathrm{F}_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{(x-\mu)^2}{\nu\sigma^2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \tag{14}$$

where Γ is the so called gamma function and ${}_2\mathrm{F}_1$ is the hypergeometric function (see Johnson et al [6]). Cauchy distribution (also called Lorentz distribution or Breit-Wigner distribution) is a special case of t-distribution when $\nu = 1$. It has two parameters: location μ and scale γ. However, its expected value and variance are not defined. Its PDF is

$$\mathrm{f}(x,\mu,\gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-\mu}{\gamma}\right)^2\right]}, \tag{15}$$

and the corresponding CDF is

$$\mathrm{F}(x,\mu,\gamma) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\gamma}\right) + \frac{1}{2}. \tag{16}$$

5 Case Study: Cryptocurrencies

A cryptocurrency is a virtual asset designed to act as a traditional currency. As an asset cryprocurrencies are priced in other media of exchange and their prices exhibit much higher volatility than those of traditional assets. This feature makes Voigt distribution a potentially suitable candidate which can be used for modeling the extreme heavy tails of distribution of returns of cryptocurrencies.

Characteristic	Bitcoin	Litecoin	XRP	Ethereum
Mean	1.6775×10^{-3}	1.2891×10^{-3}	1.8780×10^{-3}	2.9700×10^{-3}
Median	1.8925×10^{-3}	0	-2.8754×10^{-3}	-9.2067×10^{-4}
Mode	0	0	0	0
Minimum	-2.6620×10^{-1}	-5.1393×10^{-1}	-6.1627×10^{-1}	-1.3021
Maximum	3.5745×10^{-1}	8.2897×10^{-1}	1.0274	4.1234×10^{-1}
Std deviation	4.3131×10^{-2}	6.6333×10^{-2}	7.5727×10^{-2}	7.5628×10^{-2}
Skewness	-1.7169×10^{-1}	1.7429	2.0365	-3.4122
Kurtosis	11.037	28.210	31.189	70.192
Obs.	2189	2189	2091	1358

Table 1 Descriptive statistics of logarithmic return series

There are thousands of crytocurrencies in circulation. Four of them are chosen for our analysis. They are Bitcoin, Litecoin, Ripple, and Ethereum, four of five of the most liquid cryptocurrencies. The market capitalization of Bitcoin, the most liquid one, at this moment is around one hundred billion USD. Daily data on their prices is taken from CoinMarketCap provider. The series of the first three are from period 2013 - 2019, the last one is from 2015 - 2019. All price series are transformed into series of logarithmic returns. Their descriptive statistics are shown in Table 1. From Table 1 one can observe that all returns have positive mean and much higher standard errors and kurtosis compared to those of traditional returns.

Currency	Parameter	Value	Std. error	p-value
Bitcoin	μ	2.2026×10^{-3}	3.3378×10^{-4}	4.1382×10^{-11}
	σ	5.5563×10^{-3}	3.1186×10^{-3}	0.0748
	γ	1.3842×10^{-2}	9.3756×10^{-4}	2.4272×10^{-49}
Litecoin	μ	-1.2484×10^{-3}	1.4321×10^{-5}	0
	σ	2.3248×10^{-3}	5.0520×10^{-3}	0.6454
	γ	1.8983×10^{-2}	8.3686×10^{-4}	6.5933×10^{-114}
XRP	μ	-2.7425×10^{-3}	6.7597×10^{-4}	4.9680×10^{-5}
	σ	8.0916×10^{-3}	3.2924×10^{-3}	1.3983×10^{-2}
	γ	1.9827×10^{-2}	1.1407×10^{-3}	1.1859×10^{-67}
Ethereum	μ	-6.7001×10^{-4}	3.9077×10^{-3}	0.8639
	σ	2.4784×10^{-2}	3.0120×10^{-3}	2.2204×10^{-16}
	γ	1.8046×10^{-2}	1.6201×10^{-3}	8.4519×10^{-29}

Table 2 Estimated values of parameters of Voigt distribution

The return series are used to estimate the parameters of Voigt distribution by the technique described in previous sections. The estimation procedure is very stable and reliable. It always delivers the same estimates from any initial values. The estimation results are shown in Table 2. Besides the estimates, their corresponding asymptotic standard errors and p-values are also reported. For a comparison purpose, we also estimate the parameters of Cauchy distribution (values of parameters of normal distribution can be found in descriptive statistics with nonasymptotic standard deviations) and Student t-distribution using the same data series also with the MLE technique (see Balmer et al [2]). The estimation results are displayed in Table 3 in a similar fashion.

The results in Table 2 show that except the case of Ethereum, estimated value of γ is always greater than the value of σ indicating the dominance of Cauchy distribution which captures the fat tail property of returns. Moreover, as the p-values show, in the case of Bitcoin and Litecoin, the estimates of σ are statistically insignificant,

Distribution	Currency	Parameter	Value	Std. error	p-value
Cauchy	Bitcoin	δ	2.038×10^{-3}	8.273×10^{-4}	1.376×10^{-2}
		γ	1.466×10^{-2}	3.965×10^{-4}	2.9802×10^{-299}
	Litecoin	δ	-1.274×10^{-3}	5.140×10^{-4}	1.318×10^{-2}
		γ	1.911×10^{-2}	5.392×10^{-4}	3.2403×10^{-275}
	XRP	δ	-2.828×10^{-3}	6.543×10^{-4}	1.545×10^{-5}
		γ	2.109×10^{-2}	6.476×10^{-4}	9.5400×10^{-233}
	Ethereum	δ	-2.050×10^{-3}	9.356×10^{-4}	2.844×10^{-2}
		γ	2.592×10^{-2}	1.209×10^{-3}	6.8688×10^{-102}
Student t	Bitcoin	μ	2.333×10^{-3}	5.248×10^{-4}	8.752×10^{-6}
		σ	1.914×10^{-2}	6.878×10^{-4}	1.7929×10^{-170}
		ν	1.784	1.055×10^{-1}	3.5466×10^{-64}
	Litecoin	μ	-1.225×10^{-3}	6.759×10^{-4}	7.005×10^{-2}
		σ	2.388×10^{-2}	1.455×10^{-3}	1.5795×10^{-60}
		ν	1.592	1.798×10^{-1}	7.8434×10^{-19}
	XRP	μ	-2.657×10^{-3}	7.424×10^{-4}	3.441×10^{-4}
		σ	2.594×10^{-2}	7.149×10^{-4}	3.5263×10^{-288}
		ν	1.555	1.820×10^{-2}	0
	Ethereum	μ	-7.391×10^{-4}	1.103×10^{-3}	0.5029
		σ	3.562×10^{-2}	1.700×10^{-3}	1.6923×10^{-97}
		ν	2.162	2.082×10^{-1}	3.0870×10^{-25}

Table 3 Estimated values of parameters of Cauchy and Student t-distributions

meaning that the distribution of returns is very close to the Cauchy distribution. The closeness is confirmed by estimation results in Tables 2 and 3. Regarding other estimation results, the estimated values of t-distribution in Table 3 show that, again, except the case of Ethereum, in the other cases the values of ν, i.e. the number of degrees of freedom, is lower that 2, indicating the extremely heavy tails of returns of cryptocurrencies.

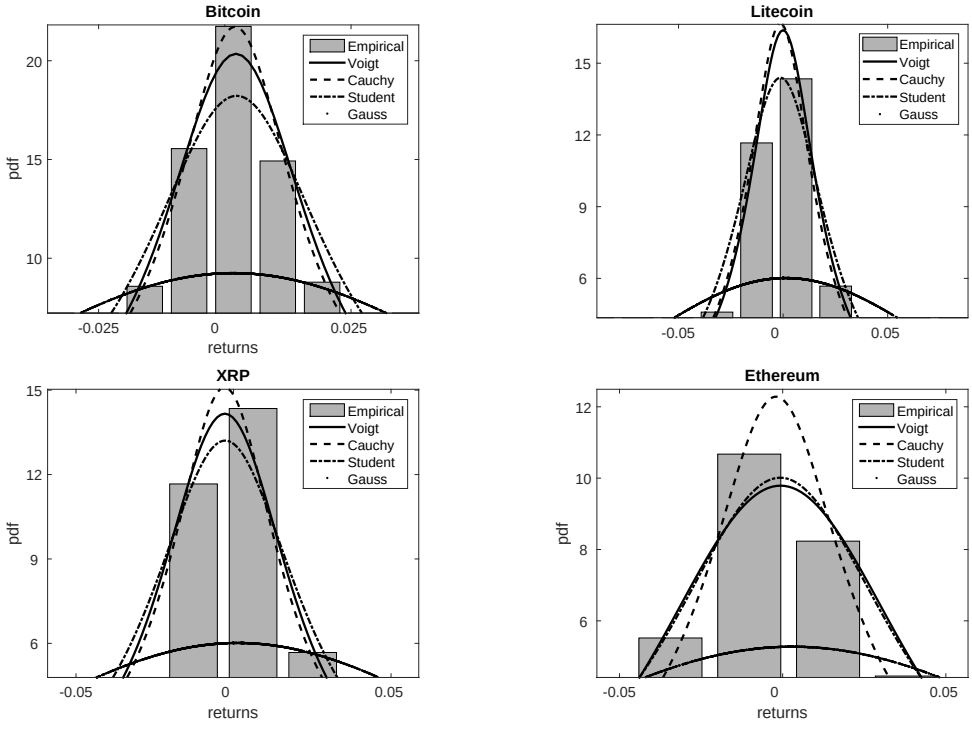


Figure 2 The PDFs of returns of Bitcoin and Litecoin

Currency	Gaussian	Cauchy	Student	Voigt
Bitcoin	3.78×10^3	4.158×10^3	4.22×10^3	4.159×10^3
Litecoin	2.83×10^3	3.523×10^3	3.57×10^3	3.523×10^3
XRP	2.43×10^3	3.158×10^3	3.20×10^3	3.159×10^3
Ethereum	1.58×10^3	1.859×10^3	1.91×10^3	1.871×10^3

Table 4 Values of log-likelihood functions at estimated optimal values of parameters

We use the estimated values of parameters of each distribution to generate the values of the corresponding theoretical PDF. These values together with the values of the empirical PDF for each series are shown in Figure 2. We display only the area of all distributions around the peak which is zoomed to show where each distribution stands. From these figures we can observe the fact that in the convolution Cauchy distribution is the dominant component except for the returns of Ethereum. Regarding t-distribution, it seems to be acapable of modeling returns of cryptocurrencies very well, similar to the results reported by Blattberg and Gonedes [4].

We also evaluate the values of log-likelihood functions used to estimate parameters of distributions using MLE. The results are put in Table 4. The likelihood values support the inferences we obtain from the results of the estimation process. By LR test, the statistically significant difference between Gaussian and Student t-distributions, between Gaussian and Voigt distributions, and Cauchy and Student t-distributions can be unequivocally inferred. Otherwise, when the estimates of σ are insignificant, the likelihood values with Voigt distribution are higher than those with Cauchy distribution. Those values with t-distribution are the highest ones indicating that t-distribution might be a better candidate for modeling the returns of cryptocurrencies.

6 Conclusion

We have investigated the ability of Voigt distribution as an alternative for modeling extremely heavy tail property of financial data. Despite the existence of the close form of its PDF, the parameter MLE procedure using directly the PDF is too time-consuming. Therefore, we have proposed the MLE procedure with FFT. This approach has been tested on simulated data. It is almost twelve times faster than using analytic PDF with reasonable accuracy. Then Voigt distribution is used to model the returns of four most liquid cryptocurrencies: Bitcoin, Litecoin, Ripple, and Ethereum. The results we obtained show our method can be used for estimation of parameters of Voigt distribution. The method is numerically stable and it provides the results comparable with those of other fat tail distributions. For the chosen series of returns of cryptocurrenices, the Voigt distribution can be a good candidate for modeling extremely fat tail property in some cases. However, to more competently answer the question of its suitability for this task, further research is needed.

Acknowledgements

This paper is financially supported by grant SGS17/196/OKH4/3T/14.

References

[1] Armstrong, B. H. (1967). Spectrum line profiles: the Voigt function. *J. Quant. Spectrosc. Radiat. Transfer* 7, 61 - 88.

[2] Balmer, D.W., Boulton, M., Sack, R.A. (1974). Optimal solutions in parameter estimation problems for the Cauchy distribution. *J. Amer. Statist. Assoc.* 69, 238 – 242.

[3] Brigham, E. O. (2002). *The Fast Fourier Transform*. New York, USA: Prentice-Hall.

[4] Blattberg, R. C., Gonedes, N. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business* 47, 244–280.

[5] Eberlein, E., and U. Keller. 1995. Hyberbolic distributions in finance. *Bernoulli* 1 (3): 281—299.

[6] Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions* Volume 1. New York: Wiley & Sons.

[7] Nolan, J. P. (2003). Modeling Financial Data with Stable Distributions. In Rachev, S. T. (editor): *Handbook of Heavy Tailed Distributions in Finance* 2. Amsterdam: North Holland, 105 - 130.

[8] Theodossiou, P. (1998). Financial Data and the Skewed Generalized t Distribution. *Management Science* 44, 1650-1661.

Portfolio Selection Model based on Drawdown Risk Measure with Different Inputs

Juraj Pekár[1], Ivan Brezina[2], Marian Reiff[3]

Abstract. In addition to assets selection based on expected returns and financial indicators, the Investor in the decision-making process on investment strategy faces the question of how to implement portfolio diversification. Portfolio theory deals with this issue, with aims to find a mix of assets with the optimization models of expected return and risk criteria. Generally, these models result in a set of effective portfolios, from which the decision maker, based on additional information, selects the best alternative. The paper deals with the model in which the risk of Drawdown is used. The starting point is known model from the literature with assumptions of known cumulative yield values and the maximum possible loss of capital for the constructed portfolio. This means that the decision maker implements a portfolio based on the yield maximization criterion, with a known maximum possible loss of capital. However, the required minimum yield value may also be known in practice. Based on this idea, a modification of the model is constructed in which, unlike the original model, the input that represents the value of the minimum yield is altered and the aim is to minimize the risk.

Keywords: risk measure, portfolio selection model, Drawdown

JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

Nowadays, it is common practice that the investors split the investments on financial markets. Financial assets are the subject of financial market trading, and the financial market is a set of institutions and instruments that provide the movement of money and capital in different forms between different economic operators based on supply and demand. Three indicators are particularly important for capital procurement for the demand on the financial market. The revenue is an indicator representing all the income an investor earns (investment appreciation). The liquidity is an indicator of the timeframe of how quickly the investment will bring cash. The risk is the probability that the money invested will not yield the expected return. The investor creates his own portfolio of assets, whose main characteristic is its efficiency. Portfolio efficiency (performance) can be characterized by the amount of return owner benefits. The profitability of the portfolio is determined by the profitability of the individual assets constituting the portfolio and the amount of capital invested in the portfolio (the average of the return on individual assets, the weights being given by the size of the invested capital).

To support decision making several calculations and analyses are made. The goal is to answer the question, how much can the investor lost in the selected period. Drawdown is the risk measure and the measure of investment strategy success, interested reader can see [1], [2], [3], [4] and [5]. On investment market Drawdown is defined as the biggest decline of portfolio return.

The maximum value of Drawdown measures highest percentage decline in portfolio value in selected period. This value can be used as the measure of the portfolio risk. It is the indicator of the portfolio management quality. The value of Drawdown depends on the maximum and minimum value of asset in the test period. It gives the investor information about investment strategy success, about the investment strategy risk and about the finance, which investor needs to realize the investment. The value of Drawdown is the indicator of portfolio risk and investment strategy success. The low value of Drawdown indicates good portfolio management and small risk of selected portfolio.

[1] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail:, juraj.pekar@euba.sk.
[2] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: ivan.brezina@euba.sk.
[3] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: marian.reiff@euba.sk.

In addition to the underlying risk rate of Drawdown, there are also derived risk rates, Maximum Drawdown, Average Drawdown, and Conditional Drawdown. The paper presents ways of calculating the stated risk rates. Based on conditional Drawdown, portfolio selection optimization models are formulated. Portfolio optimization is presented in discrete terms on presented models and its goal in the first model is defined as the maximum average portfolio return at the end of the reporting period, while the second model is designed to minimize risk based on the Drawdown risk. Part of the paper is the formulation of a modified model in GAMS software.

2 Risk measure based on Drawdown and expected return

As mentioned above, Drawdown is defined as the biggest decline in portfolio return. In this section, we will describe the input data procedure of the model (calculation of the cumulative yield vector and how to obtain the expected return on the asset) and then define the Drawdown and the derived risk rates (Maximum Drawdown, Average Drawdown and Conditional Drawdown) that are used to formulate our model.

Let consider n assets and let P_{jt} be the price of j-th asset, j = 1, 2,...n, in the time t =1, 2,...T, where T is the length of the reference period. Consequently, the rate of cumulated yield of j-th asset, j = 1, 2,...n, in the time t is possible to expressed as

$$y_{jt} = \frac{P_{jt} - P_{j1}}{P_{j1}} \tag{1}$$

Expected return of j-th asset, j = 1, 2,...n, based on the cumulative yield value:

$$E_j = \left(1 + y_{jT}\right)^{\frac{1}{T}} - 1 \tag{2}$$

The risk rate of Drawdown (*DD)* at the considered time is defined as the difference between the maximum value of the asset (portfolio) in previous periods and its present value at the time t, which takes the values from the analyzed reference period T. Cumulative yields of the j-th asset, j = 1, 2,...n, in each period t = 1, 2,...T are denoted as y_{jt}. The value $DD_j(t)$ of j-th asset (portfolio), j = 1, 2,...n, in time t, t =1, 2,...T can be calculated as the difference between the maximum value in v-th period (v = 1,2 ,...t) and the value in time t. According to [2] it is possible to calculate as:

$$DD_j(t) = \max_{v=1,2,...t} y_{jv} - y_{jt} \tag{3}$$

In addition to the indicator under consideration, we will also deal with the derived risk rate [2], maximum drawdown (MDD). Because it is natural that the investor is particularly interested in the likelihood that an investment in an asset will not reach the expected return, it is precisely the risk rate of the maximum drawdown that allows you to track the greatest downfall in asset value. This indicator is based on historical records and describes the largest decrease in value that occurred during the analyzed period (T) Maximum Drawdown is the largest cumulative percentage decline in the Net Asset Value of portfolio from the highest or peak value to the lowest or trough value after the peak. Maximum Drawdown $MDD_j(T)$ for j-th asset, j = 1, 2,...n, is possible to calculate as:

$$MDD_j(T) = \max_{t=1,2,...T} DD_j(t) \tag{4}$$

Average Drawdown (AvDD) during the analyzed period (T) for j-th asset, j = 1, 2,...n, can be calculated as (the Drawdowns of each year are averaged to come up with an average annual Drawdown):

$$AvDD_j(T) = \frac{1}{T}\sum_{t=1}^{T} DD_j(t) \tag{5}$$

For the data in the observed time interval t = 1, 2,...T according to [2] the Conditional Drawdown (CDaR) can be defined for j-th asset, j = 1, 2,...n, as:

$$CDaR_j(T)_\alpha = \min_{DaR}\left\{ DaR + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T}\left[DD_j(t) - DaR\right]^+ \right\} \tag{6}$$

Where DaR is the Drawdown (DD) threshold of the portfolio, this value can only be exceeded by$(1-\alpha)T$ observations. Notation + in formula (6) means, only the positive values are taken into account.

3 Portfolio Drawdown optimization

Portfolio Drawdown (DD) optimization was proposed by Cheklov et al. [2]. The aim is to find the optimal portfolio allocation, taking into account DD. The aforementioned authors, in optimizing the DD portfolios, include risk functions among the constraints, while setting a maximum average sub-threshold of the investment fall. Portfolio optimization is then based on revenue maximization. Let us further consider that **u** represents a vector (u_0, u_1,...u_T) of free variables, where u_t value represents the maximum cumulated yield value over time 0, 1,...T. Let value a represents maximum average sub-threshold of the investment fall loss value of the portfolio. Parameter values E_j (j = 1, 2,...n) denote the expected return on each asset. Let variable DaR represents the threshold for the DD portfolio. Furthermore, let variable z_t represent the values of DD subthreshold based on the significance level α. In addition to the variables listed, variables w_1, w_2,... w_n, denote values of the individual asset weights. Then, the expected portfolio return maximization task with regard to the limits of the maximum allowable loss of capital can be formulated according to [2] as:

$$\max E\left(w_1, w_2, ...w_n, u_0, u_1, ...u_T, z_1, z_2, ...z_T, DaR\right) = \sum_{j=1}^{n} w_j E_j \tag{7}$$

$$DaR + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T} z_t \le a, t = 1, 2, ...T \tag{8}$$

$$z_t \ge u_t - \sum_{j=1}^{n} y_{jt} w_j - DaR, t = 1, 2, ...T \tag{9}$$

$$u_t \ge \sum_{j=1}^{n} y_{jt} w_j, t = 1, 2, ...T \tag{10}$$

$$u_t \ge u_{t-1}, t = 1, 2, ...T \tag{11}$$

$$u_0 = 0 \tag{12}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{13}$$

$$w_1, w_2, ...w_n, z_1, z_2, ...z_T \ge 0 \tag{14}$$

The aim of the paper is to modify the above mentioned task in which the portfolio optimization is based on conditional Drawdown minimization. Let consider the input data listed in the previous section, and unlike the previous model, let the expected yield value to be known. Therefore, we do not consider the entry value of the conditional Drawdown, but the value of the yield required. Let E_p be the value that represents the minimum value of the desired portfolio yield. The modified task of CDaR portfolio minimization with respect to the expected yield constraints is defined as:

$$\min CDaR\left(w_1, w_2, ...w_n, u_0, u_1, ...u_T, z_1, z_2, ...z_T, DaR\right) = DaR + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T} z_t \tag{15}$$

$$\sum_{j=1}^{n} w_j E_j \ge E_p \tag{16}$$

$$z_t \ge u_t - \sum_{j=1}^{n} y_{jt} w_j - DaR, t = 1, 2, ...T \tag{17}$$

$$u_t \ge \sum_{j=1}^{n} y_{jt} w_j, t = 1, 2, ...T \tag{18}$$

$$u_t \ge u_{t-1}, t = 1, 2, ...T \tag{19}$$

$$u_0 = 0 \tag{20}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{21}$$

$$w_1, w_2, ...w_n, z_1, z_2, ...z_T \ge 0 \tag{22}$$

We formulate the problem in software using linear model (15) – (22). Model is implemented in the software GAMS (solver CONOPT) on PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8 (Figure 1).

```
Sets t
sub(t)
i;
Scalar Ep , alfa , TT ;
Parameters E(i);
Table y(sub,i)    ;
Variables CDaR,u(t),a,DaR;
Positive variables w(i),z(sub);
Equations
Objective
Constrain
Constrain0
Constrain1(sub)
Constrain2(sub)
Constrain3(sub)
Constrain4
Constrain5;
Constrain. sum(i,w(i))=e=1;
Constrain0..Ep=g=sum((i),w(i)*E(i));
Constrain1(sub)..z(sub)=g=u(sub)-sum(i,y(sub,i)*w(i))-DaR;
Constrain2(sub)..u(sub)=g=sum(i,y(sub,i)*w(i));
Constrain3(sub)..u(sub)=g=u(sub-1);
Constrain4..u('0')=e=0;
Constrain5..a=g=DaR+sum(sub,z(sub))/((1-alfa)*TT);
Objective. CDaR=e=a;
Model CDaR1 /all/
Solve CDaR1 using lp minimazing CDaR;
```

Figure 1 GAMS code

The result is the solution of this problem, which consist of asset proportions of invested capital value in the optimal portfolio.

4 Conclusion

The paper deals with the modification of the conditional Drawdown based portfolio selection model in which, unlike the original model [2], we do not consider the conditional Drawdown as input parameter but the desired portfolio yield value. In addition to the aforementioned change, a different approach was used to determine the value of the expected return on the individual assets. The geometric average was used unlike in the original paper, it better represents its expected value. Another modification is the definition of the variable domain, which represents the weights of investing in individual assets, and we expect to invest all available funds in the calculated portfolio. The proposed modification enables to determine the required values of input parameters in percentages resulting in better interpretation of obtained results.

The problem solution in GAMS programming environment (CONOPT solver) documents the possibility of calculating the optimal solution that is the basis for the investor's decision. Of course, alternative program products can also be used to calculate.

Additional knowledge of optimal portfolio allocation is one of the benefits of professional management for the right investment decisions. The fact that the return on plan assets are liable to fluctuate greatly, affect a suitable investment strategy. Practical Conclusions: Understanding the relationship between risk and return well enough to tailor an investment strategy to the individual needs of a client does not require knowledge of complex theoretical constructs. From a practical point of view, it may be more appropriate for the investor to use a modified model because, in the original model, it is problematic to determine the CDaR value, which in the original model represents the input parameter and its determination is complicated by different asset prices. Unlike the original model, in the proposed model the value of the expected yield that the investor can clearly determine serves as the input parameter in the presented model. The presented approach represents an interesting modification, which can be more acceptable and understandable on the basis of transparency and simpler economic interpretation of input parameters for the investor.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. No. 1/0351/17 Application of selected models of game theory to solve some economic problems of Slovakia

References

[1] Brezina, I., Pekár, J., Brezina, I. jr. (2018). Manažment portfólia na základe miery výkonnosti Drawdown. In *Competition: Proceedings of 10th Annual International Scientific Conference*, 17th–18th May, 2018, Jihlava: College of Polytechnics Jihlava.

[2] Cheklov, A., Uryasev, S. and Zabarankin, M. (2003). Portfolio Optimization with Drawdown Constraint. In: Scherer B (Ed.): *Asset and Liability Management Tools* (pp. 263–278). London: Risk Books.

[3] Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling Drawdowns. *Mathematical Finance*, 3, 241–276 .

[4] Pekár, J., Brezina, I., Brezina, I. jr. (2018). Portfolio Selection Model Based on Drawdown Performance Measure. In *International Conference. Mathematical Methods in Economics - MME 2018: Proceedings from the 36th International Conference*, Jindřichuv Hradec, Czech republic, September 12-14, 2018. Praha: Matfyzpress.

[5] Pekár, J., Čičková, Z. and Brezina, I. (2016). Portfolio performance measurement using differential evolution. *Central European Journal of Operations Research*, 24, 421-433.

Labour market frictions and vacancies: small open economy DSGE model

Jakub Bechný[1], Osvald Vašíček[2]

Abstract. The goal of this paper is to incorporate job vacancies into a small open economy dynamic stochastic general equilibrium model with involuntary unemployment. The observed vacancies are introduced into the model by using a variant of matching function. The model is estimated by using a Bayesian approach and 16 observed variables from 2001Q1 to 2018Q4 for the Czech economy. The observed vacancies allow us to identify hiring cost shock which has a moderate impact on inflation. However, the observed vacancies do not improve the overall predictive ability of the model, since the model over-predicts the vacancies' volatility.

Keywords: Bayesian estimation, DSGE model, in-sample forecasts, job vacancies, labour market, predictive ability, small open economy.

JEL Classification: C32,E17
AMS Classification: 91B51

1 Introduction

The Czech labour market exhibits a remarkable development. It has recently gone through a turbulent period of the Great Recession, followed by the 2012-2013 crisis ended up with the exchange rate commitment adopted by the Czech National Bank. The unemployment rate reached almost 8 % during the Great Recession, but the situation is completely different right now. The Czech labour market is extremely tight and exhibits symptoms of the labour shortage. The unemployment rate has reached 2 % by the end of 2018 and is currently the lowest in the European Union. There were less than 120 000 unemployed persons and the job vacancies at the same time exceeded 320 000. Furthermore, there is a possible mismatch between the skills offered by the unemployed and the demands of the employers. However, standard larger-scale DSGE models usually do not explicitly model involuntary unemployment, and incorporation of job vacancies is even rarer.

Consequently, our goal is to incorporate job vacancies into a small open economy dynamic stochastic general equilibrium (DSGE) model with real wage rigidity and involuntary unemployment. We use our modification of Sheen & Wang [11] model. The observed vacancies are linked with the marginal costs of intermediate firms (the key determinant of the inflation in the New Keynesian models) by using a variant of matching function proposed by Blanchard & Galí [4]. We use 16 observed variables from 2001Q1 to 2018Q4 for the Bayesian estimation of the model. The observed growth of vacancies allows us to identify the hiring cost shock, which has a moderate impact on the inflation rate (0.3 percentage points in response to 1 sd. shock). However, in-sample conditional forecasts point to a higher predictive ability of the model estimated without observed vacancies. The observed vacancies improve predictive ability only for inflation, nominal interest rate, and unemployment.

The research on the Czech labour market by using DSGE approach and observed vacancies has been limited. The exceptions are Němec [9], Pápai [10], and Chalmovianský [6], who also use a matching function approach. We contribute to this literature by exploring the benefits of the vacancies as an additional observable variable for the model. However, our model is more complex compared with the previous authors, allowing for more detailed analysis of the economy. Secondly, we do not pre-filter the observed data by HP filter which leads to loss of information, but we link the model variables with the observed growth rates of the data. The remainder of this paper is organised as follows. Section 2 sketches the structure of our model, results are discussed in Section 3.

2 Model

We employ a modified version of a small open economy model with real wage rigidity, originally proposed by Sheen & Wang [11]. The model contains a standard New Keynesian closed economy of Christiano et al. [7], the

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, jakub.bechny@gmail.com
[2] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, osvald.vasicek@econ.muni.cz

small open economy structure of Adolfson et al. [1], and explicitly modelled involuntary unemployment using the Nash-Bargaining and the real wage rigidity of Hall [8]. Our model contains a unit-root technology shock which is a source of a balanced growth path growth of all real variables.

There is one type of representative household that attains utility from leisure, real balances and a basket of domestically produced and imported consumption goods, subject to habit formation. The representative household accumulates physical capital investing a basket of domestically produced and imported investment goods. The central bank follows a standard Taylor rule and responds to the inflation rate, the output, and the real exchange rate. The world economy is proxied by the euro area data and is exogenous to the domestic economy. Our variant of the model contains five types of representative intermediate firms – domestic good producers, consumption importers, investment importers, importers of re-exported good, and exporters. Each intermediate firm sets its price based on a markup over its real marginal costs, and the model thus contains five New Keynesian Phillips curves. Services of capital and labour are used only by producers of domestic good, depending on the relative wage and rental rate of capital.

We introduced several modifications into the original Sheen and Wang model which were discussed in Bechný & Vašíček [2]. In this paper, we introduce the job vacancies into the model by using a variant of the matching function, as is suggested by Blanchard & Galí [4]

$$\frac{V_t}{H_t} = \varepsilon_t B x_t^{\vartheta} \zeta_t^x \equiv g_t \tag{1}$$

where V_t is number of vacancies, H_t denotes hiring, x_t is labour market tightness with elasticity ϑ, B affects steady-state hiring costs, ε_t is temporary technology shock, and ζ_t^x is hiring cost shock. The hiring costs g_t then play a role during Nash-bargaining about the real wage, and also affect marginal costs of the domestic intermediate firm – one of the key determinants of the inflation dynamics in the New Keynesian models.

In the empirical part of the paper, we estimate two variants of the model – with and without the growth rate of vacancies in the set of observed variables. The observed vacancies allow us to identify the hiring cost shock ζ_t^x; variance of this shock is restricted to zero in the variant of the model without observed vacancies. We investigate whether information from observed vacancies improved the predictive ability of the model, and analyse impulse responses to the hiring cost shock.

3 Results

All equilibrium conditions of our model were stationarised by a unit root technology and log-linearised. The whole system of eighty-four equations was solved numerically in Matlab toolbox Dynare, cast into a state-space form, and its parameters were estimated by using a standard Metropolis-Hastings algorithm. To deal with possible non-stationarities in the data, we employed diffuse Kalman filter and a Monte-Carlo based optimisation routine to initialize the Metropolis-Hastings. We use quarterly data for the Czech economy from 2001Q1 to 2018Q4, taken from the Czech National Bank database: the real output, consumption, investment, imports, exports, wages, CPI inflation, GDP deflator inflation, nominal interest rate (PRIBOR), inflation target, CZK/EUR exchange rate, unemployment rate, and for one variant of the model also job vacancies taken from the OECD database. The foreign economy is proxied by the euro area real output, CPI inflation, and nominal interest rate (EURIBOR). We use the growth rates of output, its components, and of wages and vacancies as the observed variables. The growth rates, the inflations and the interest rates are stationarised by demeaning, and we stationarise the unemployment rate by the NAIRU taken from the OECD database.

The first potential benefit of the vacancies among the observed variables is in a fact that they allow us to identify one additional structural shock – the hiring cost shock ζ_t^x. Figure 1 shows impulse responses of selected variables to this shock. The shock of 1 sd. induces an initial drop in the hiring rate of almost 6 percentage points (p.p.), and an increase in the vacancies posting of roughly the same magnitude. The wages drop almost by 0.15 %, and the unemployment increases 0.1 p.p. above the NAIRU. The real exchange rate slightly appreciates, and the marginal costs increase almost by 0.1 %, which induces a moderate inflationary pressure of 0.3 p.p.. The hiring cost shock also reduces the real output growth by 0.3 p.p.. Note that in Bechný & Vašíček [3] we showed that without observed vacancies, the hiring cost shock has no impact on the growth and inflation. Incorporation of the vacancies as the observed variable thus allows for identification of impact of the hiring cost shock on the economy.

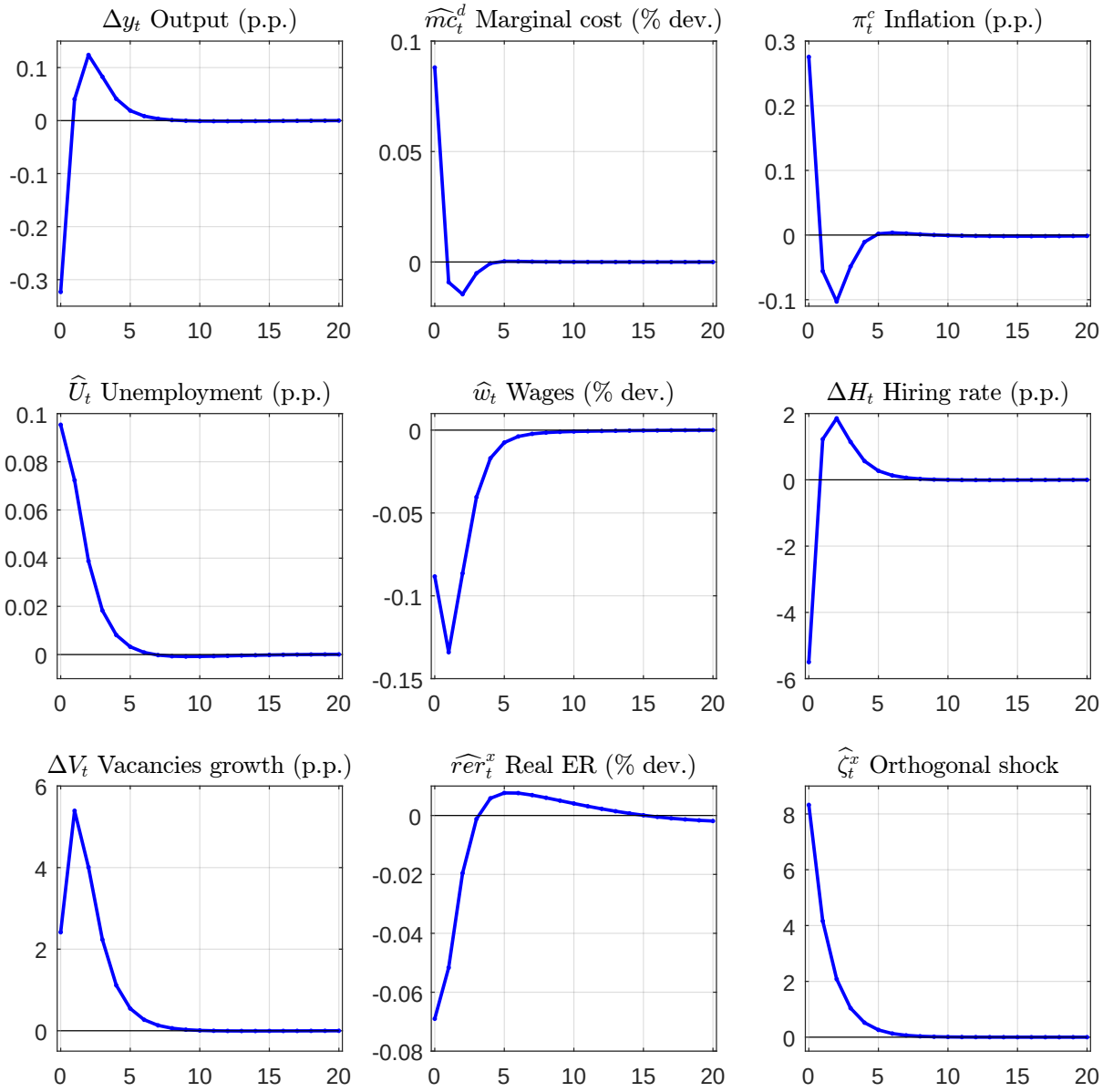


Figure 1 Impulse response functions to hiring cost shock

Notes: Output, inflation, hiring rate and vacancies are measured as growth in percentage points; output and inflation rate are annualised. Marginal cost, wages, real exchange rate and hiring cost shock are expressed in percentage deviations from steady state. The unemployment rate is measured as a deviation from the NAIRU in percentage points.

Furthermore, we evaluate the predictive ability of the model estimated with and without vacancies as the observed variable using the root mean squared errors (RMSEs) of in-sample conditional forecasts at a horizon of 1-8 quarters. The predictive efficiency cascades for selected variables are for illustration visualised in Figure 2. The results in Table 1 show unambiguously better predictive ability of the model without observed vacancies for output, imports, and exports. The model without observed vacancies provides generally better forecasts also for consumption, investments, wage, and CPI inflation. On the other hand, the information from the observed vacancies improves the predictive ability for GDP deflator inflation, nominal interest rate, and unemployment.

Based on the results of the in-sample conditional forecasts we conclude that information from the observed vacancies did not improve the overall predictive ability of the model. The main problem is that our matching function modelling approach has problems with the accuracy of the vacancies' predictions. This illustrate quite high RMSEs in Table 1 and imprecise predictive efficiency cascades for vacancies in Figure 2. The low accuracy of the vacancies' forecasts is caused by the high volatility of the observed vacancies' growth - to capture this, our model needs a volatile hiring cost shocks. Those volatile shocks then decrease the accuracy of forecasts during our

in-sample simulations.

Horizon	Output		Consumption		Investments		Imports	
(quarters)	Unobs.	Obs.	Unobs.	Obs.	Unobs.	Obs.	Unobs.	Obs.
1	**1.0963**	1.1251	0.85423	**0.85069**	2.6058	**2.5997**	**2.4067**	2.4878
2	**1.2633**	1.2638	0.90086	**0.896**	**2.9663**	2.9762	**2.6784**	2.7696
3	**1.2494**	1.2565	**0.94125**	0.94612	**2.9642**	2.9669	**2.7734**	2.8424
4	**1.2432**	1.2599	**0.95263**	0.95876	2.944	**2.9366**	**2.7548**	2.8527
5	**1.2184**	1.2401	**0.91297**	0.9198	**2.7727**	2.7741	**2.6263**	2.7346
6	**1.2159**	1.2355	**0.88318**	0.89082	**2.3831**	2.4033	**2.6461**	2.7398
7	**1.2171**	1.2403	**0.87204**	0.87382	**2.3681**	2.3766	**2.6365**	2.7313
8	**1.2061**	1.2292	**0.83375**	0.83663	**2.4279**	2.4363	-	-
Horizon	Exports		Wage		CPI inflation		Deflator inflation	
(quarters)	Unobs.	Obs.	Unobs.	Obs.	Unobs.	Obs.	Unobs.	Obs.
1	**2.3756**	2.5321	**1.4355**	1.5038	**1.0249**	1.0581	**1.0013**	1.0062
2	**2.4063**	2.5786	1.5732	**1.5711**	**1.3121**	1.3279	**1.2123**	1.2399
3	**2.4778**	2.5933	**1.4984**	1.5073	1.3412	**1.3221**	1.273	**1.2428**
4	**2.4894**	2.6503	**1.4933**	1.5035	1.2476	**1.2287**	1.1683	**1.1418**
5	**2.4376**	2.6031	**1.4917**	1.5091	1.1259	**1.1226**	1.0654	**1.0504**
6	**2.4723**	2.623	**1.4865**	1.4992	**1.0769**	1.0852	0.97686	**0.96847**
7	**2.4572**	2.6237	**1.465**	1.4818	**1.0438**	1.0606	0.9205	**0.91639**
8	**2.4456**	2.6245	**1.4585**	1.4761	**1.0452**	1.0622	0.91777	**0.91612**
Horizon	Interest rate		Unemployment		Vacancies			
(quarters)	Unobs.	Obs.	Unobs.	Obs.	Unobs.	Obs.		
1	**0.40401**	0.42573	**0.50065**	0.52566	-	31.3528		
2	**0.96924**	0.98044	**0.94754**	0.96416	-	30.5928		
3	1.5069	**1.483**	1.2421	**1.236**	-	25.4693		
4	1.9254	**1.8663**	1.4038	**1.3881**	-	24.0201		
5	2.1891	**2.1016**	1.4891	**1.4741**	-	23.8796		
6	2.3421	**2.2339**	1.5247	**1.5122**	-	24.2975		
7	2.429	**2.3127**	1.5325	**1.5217**	-	24.4552		
8	2.4656	**2.3516**	1.5185	**1.5054**	-	24.643		

Table 1 Root mean squared errors of in-sample conditional forecasts from 2006Q1 onwards
Notes: The table compares RMSE at 1-8 quarters horizon of the variant of the model estimated with (Obs.) and without (Unobs.) vacancies as the observed variable. The bold number denotes better predictive ability for given variable and horizon. The forecasts are computed conditionally to the known the future paths of the foreign output, inflation, interest rate, the exchange rate, and the inflation target.

4 Conclusions

The goal of this short paper was to incorporate job vacancies as another observed variable into a small open economy DSGE model of the Czech economy. The observed vacancies were linked with the hiring costs (and thus with the marginal costs and inflation) by using a variant of matching function proposed by Blanchard & Galí [4]. We showed that observed vacancies allow for identification of the hiring cost shock, which can induce a moderate inflationary pressure of 0.3 percentage points (in response to the shock of 1 sd.). Secondly, we compared the predictive ability of the model estimated with and without vacancies as the observed variable. The information from the observed vacancies improved the predictive ability only for inflation, unemployment, and nominal interest rate. However, the model with unobserved vacancies and without the hiring cost shock exhibited better predictive ability for the remaining seven investigated variables.

On the one hand, the model with observed vacancies exhibits quite intuitive behaviour in response to the

hiring cost shock, and the improved predictive ability for the inflation might be interesting from the monetary policymakers' point of view. On the other hand, our modelling approach to the vacancies is disputable, since the model must use a large hiring cost shocks to fit high volatility of observed vacancies. Similar problems are reported by Buss [5] who also concluded that the better fit of vacancies' volatility comes at the cost of a worse predictive ability. These results highlight possible misspecification of the modelling approach based on the matching function.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0963/2018. This support is gratefully acknowledged.

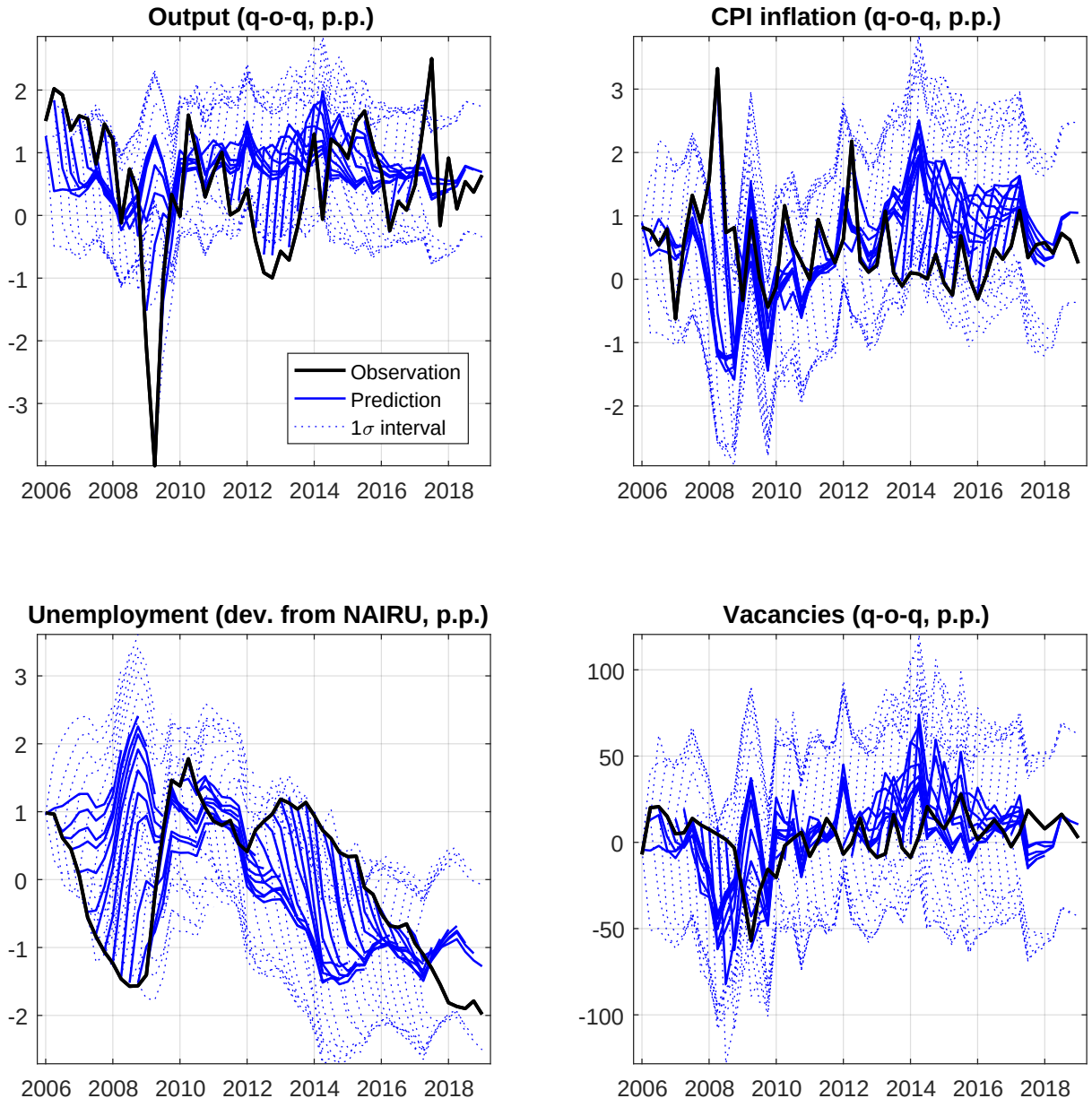


Figure 2 Predictive efficiency cascades for selected variables, variant with observed vacancies

References

[1] Adolfson, S., Laséen, J., Lindé, J. & Villani, M. (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics*, 72(2), 481–511.

[2] Bechný, J. & Vašíček, O. (2017). A Medium-Scale DSGE Model with Labour Market Frictions. In P. Pražák (Ed.), *35th International Conference Mathematical Methods in Economics, Conference Proceedings* (pp. 42–47). Hradec Králové: University of Hradec Králové.

[3] Bechný, J. & Vašíček, O. (2018). Labour market frictions and growth in small open economy: DSGE model. In Váchová, L. & Kratochvíl, V. (Eds.), *36th International Conference Mathematical Methods in Economics, Conference Proceedings* (pp. 31–31). Prague: Faculty of Management, University of Economics.

[4] Blanchard, O. J., J. & Galí, J. (2010). Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics*, 2(2), 1–30.

[5] Buss, G. (2015). *Search-and-Matching Frictions and Labour Market Dynamics in Latvia*. Working Papers 2015/04, Latvijas Banka.

[6] Chalmovianský, J. (2018). Impact of labour-market segmentation on identification of parameters in a DSGE Framework. In Reiff, M. & Gežík, P. (Eds.), *Quantitative Methods in Economics; Multiple Criteria Decision Making XIX* (pp. 24–31). Bratislava: University of Economics.

[7] Christiano, L. J., Eichenbaum, M. & Evans, Ch. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1), 1–45.

[8] Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95(1), 50–65.

[9] Němec, D. (2013) Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *Czech Economic Review*, 7(1), 21–42.

[10] Pápai, A. (2017). *Modelling labour market rigidities*. Dissertation thesis. Masaryk University, Faculty of Economics and Administration, Brno. Available from: https://is.muni.cz/th/322878/esf_d/.

[11] Sheen, J. & Wang, B. Z. (2016). Assessing labor market frictions in a small open economy. *Journal of Macroeconomics*, 48, 231–251.

Portfolio optimization with VaR and CVaR: the case of Gold and Euro portfolio

Jiří Málek [1], Quang Van Tran [2]

Abstract. The correlation between the gold price and the USD/EUR exchange rate is often negative with some exceptions in shorter periods. Hence, gold is known to be used as a useful tool to secure dollar deposits. This paper examines the portfolio optimization possibility using VaR and CVaR as risk measures. We use the t-distribution and the Normal Inverse Gaussian (NIG) distribution as well as the normal distribution as a comparative benchmark to approximate the empirical probability distribution. We construct several portfolios with various composition of gold and assets denominated in euro. First, parameters of all candidate distributions are estimated. Then, the risk measures of these portfolios are computed. We also repeat this procedure for those periods with highest positive and negative correlation. The obtained results show that it is possible to determine the optimal structure of these portfolios with VaR and CVaR as a risk measure

Keywords: dollar, gold, normal inverse Gaussian distribution, Student t-distribution, VaR, CVaR

JEL Classification: G11, C13, C40
AMS Classification: 90A09, 62G32

1 Introduction

It is a stylized fact that the gold price (in USD) is negatively correlated with the exchange rate of US dollar (USD) against major world currencies. That is why gold is often used as a security against the decline of USD exchange rate. If an investor in the Eurozone owns a dollar deposit, he will try to secure his wealth against the fall in exchange rate by purchasing gold. Since the negative correlation is changing over time and sometimes it can be even positive, we want to analyze the riskiness of this hedging strategy (gold and dollars in terms of Euro currency). For this purpose, in this paper, using daily series of gold price and exchange rate EURUSD from 1/1999 to 11/2018 we create several portfolios with different gold and currency ratios. Then the Normal Inverse Gaussian (NIG) and Student t-distributions are used to approximate the log-return distribution of these portfolios as they have been proven to be the most appropriate for approximating the empirical distributions of most marketable assets. The rest of this paper is structured as follows. In the first part, the gold price behavior is briefly described. In the second part, the basic characteristics of the NIG and t-distributions are presented. After that, in the empirical part, we perform parameter estimation of the two distributions using MLE from data for different portfolios. Then the values of VaR_{99} , $CVaR_{97.5}$ of these portfolios are calculated and compared. Based on the results of the experiment, some conclusions are inferred.

2 Dollar and gold correlation

To hedge against falling dollar, investors have to look for alternative investment possibilities to secure the value of their wealth. Weak US dollar increases the value of other countries' currencies. This increases the demand for gold and other commodities. It also increases the price of gold. Therefore, when the U.S. dollar starts to lose its value, gold is an alternative. However, the U.S. dollar and gold do not always move inversely (see Figures 1 and 2 - in Figure 1 the development of correlation coefficients between returns of gold price and EURUSD exchange rate of a window of length 65 for period 1999 - 2018 is shown, in Figure 2 the development of correlation coefficients between returns of gold price and EURUSD exchange rate of two periods is displayed. The first period is 1999 - 2000 when correlation coefficients fluctuate around zero, slightly below to be exact, and the second period of 2007 - 2009, when they seem to be most negative.) And even though they do, the size of their correlation relationship is not constant over time. For example, the traditional inverse relationship broke down during the two year period of 1978-1980. One of the reason is assumed to be that gold prices are sometimes driven upwards not by the

[1] University of Economics in Prague, Department of Banking and Insurance, nám. W. Churchilla 4, Praha 3, Czech Republic, malek@vse.cz
[2] University of Economics in Prague, Department of Monetary Economics and Policy, nám. W. Churchilla 4, Praha 3, Czech Republic, tran@vse.cz

relative weakness of the US dollar against other currencies, but instead by the outflow of paper currencies. This was probably the case of 1978-1980, when investors feared global recession following the oil crises or the collapse of the world's monetary system.

The second reason even further strengthens the role of gold as the global currency. Gold can be an insurance policy against financial crashes or even collapse of the monetary system. However, the U.S. dollar exhibits similar tail risk properties. The dollars are also seen as the world's safe haven currency. The U.S. dollar is still the main reserve currency- over half of the total amount of paper dollar stock is held outside the USA and U.S. Treasuries are busily bought during crises. It should be clear now why the U.S. dollar and gold can move up or down together. Investors may choose both as safe havens: paper dollars and gold during global catastrophes or crises in another currency. The best example may be the period from November 2008 to February 2009, when both gold and the U.S. dollar were generally rising due to the financial crisis. Such a co-movement proved that gold price behaves sometimes as a hedge against a dollar-denominated-system rather than the dollar itself. This is why gold is not only a hedge against stocks on average, but also a safe haven in extreme stock market conditions. Similar movement occurred in November 2010, when gold and the U.S. dollar rose together due to growing concerns about Ireland and the Eurozone's situation.

To sum up, a rising in gold price while the U.S. dollar remains flat means that investors are concerned about the condition of the global economy and the international monetary system, while a simultaneous rally in the U.S. dollar and gold indicates that investors worry about the global economic stability outside of the USA and thus shift capital to the two most important safe assets. Although U.S. dollar is a fiat money it is still considered as one of the best of the paper currencies. Interest rates also affect the price of gold. Gold does not yield interest in itself; therefore, it must compete with interest-bearing assets for demand. When interest rates move higher, the price of gold tends to fall, since it costs more to carry the metal. In other words, other assets will command more demand because of their interest rate component.

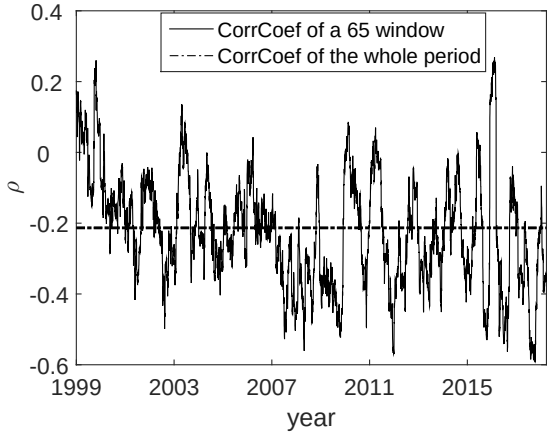


Figure 1: Correlation of returns of gold price and Euro/US Dollar exchange rate over time

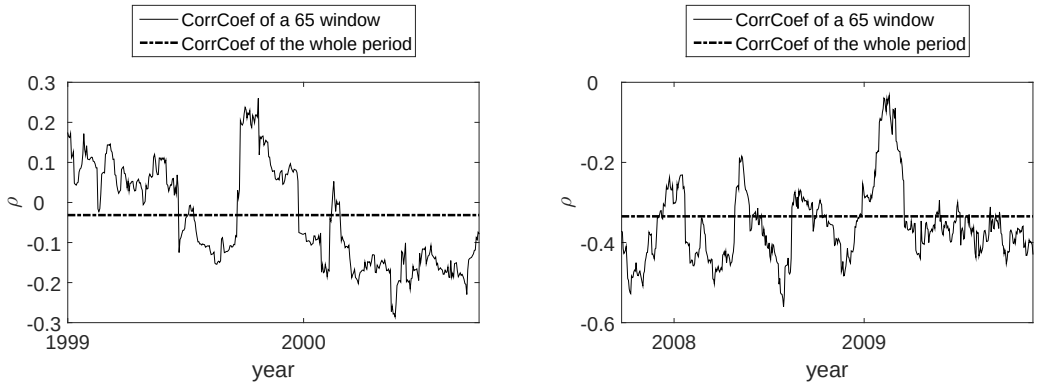


Figure 2: Correlation of returns of gold price and Euro/Dollar exchange rate in two special periods 1999–2000 and 2007–2009

3 Normal Inverse Gaussian an Student t-distributions

These two distributions are special cases of the generalized hyperbolic distribution (GHD) which is characterized by five parameters $\theta = (\lambda, \alpha, \beta, \delta, \mu)$ and its probability density function is

$$f_{\mathrm{GH}}(x;\theta) = \kappa[\delta^2 + (x-\mu)^2]^{\frac{1}{2}(\lambda-\frac{1}{2})} K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\exp(\beta(x-\mu)), \tag{1}$$

where $\kappa = \frac{(\alpha^2-\beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$, and K_λ is the modified Bessel function of the third kind with index λ. All moments of a random variable generalized hyperbolic distributed exist and the first two of them are

$$\mathbb{E}(X) = \mu + \frac{\beta\delta^2}{\zeta}\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} \tag{2}$$

$$\mathrm{Var}(X) = \delta^2\left\{\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} + \left(\frac{\beta\delta}{\zeta}\right)^2\left[\frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)}\right)^2\right]\right\} \tag{3}$$

where $\zeta = \delta\sqrt{\alpha^2-\beta^2}$. The semi-fat tail property of the GHD comes from the fact that for the generalized hyperbolic distribution: $P(X < x) \approx |x|^{\lambda-1}\exp[(\alpha+\beta)x]$ as $x \to -\infty$. When $\lambda = -\frac{1}{2}$, we get normal inverse Gaussian (NIG) distribution and its density becomes

$$f_{\mathrm{NIG}}(x;\theta) = \frac{\alpha\delta}{\pi}\exp[\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)]\frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\sqrt{\delta^2+(x-\mu)^2}}. \tag{4}$$

When $\theta = (-\frac{\lambda}{2}, 0, 0, \sqrt{\lambda}, \mu)$, we get Student t-distribution with pdf

$$\mathrm{f}(x,\mu,\sigma,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}, \tag{5}$$

where $\nu = -\frac{\lambda}{2}, \sigma = \sqrt{\lambda}$. Parameter ν is called the number of degree of freedom. The corresponding CDF is

$$\mathrm{F}(x,\mu,\sigma,\nu) = \frac{1}{2} + \frac{x-\mu}{\sigma}\Gamma\left(\frac{\nu+1}{2}\right)\frac{{}_2\mathrm{F}_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{(x-\mu)^2}{\nu\sigma^2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \tag{6}$$

where Γ is the so called gamma function and ${}_2\mathrm{F}_1$ is the hypergeometric function. In general, t-distribution has heavier tail than NIG distribution.

4 VaR and CVaR

Value at risk (VaR) is a measure which quantifies the level of risk of a portfolio, financial position or an institution for a specific time frame. Formally, for a profit-loss random variable X[1] the VaR at the level $\alpha \in (0,1)$ VaR_α is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} | F_X(x) \geq 1-\alpha\}, \tag{7}$$

where $F_X(x)$ is cumulative distribution function of X. VaR evaluates a potential loss of investment and the probability of occurrence of this loss.

The disadvantage of risk measure VaR is it is not subaditive, e.g. VaR of a portfolio may be higher than a sum of VaR of its individual components. To overcome this drawback, Conditional Value at Risk (CVaR) has been proposed as an alternative riskiness measure to VaR. CVaR is defined as

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_0^{1-\alpha}\mathrm{VaR}_s(X)ds \tag{8}$$

CVaR has gradually gained its utilisation among practitioners. Bank for International Settlements (BIS) has been pushing banks to use CVaR instead of VaR. Recommended α for CVaR is 0.975 because for normal distribution $\mathrm{VaR}_{0.99}$ and $\mathrm{CVaR}_{0.975}$ are equal. CVaR can be understood as conditional expected value under conditions that the loss exceeded the corresponding VaR.

[1] Here our notation is that profits are positive, losses are negative.

5 Empirical analysis and results

For our analysis, two series of gold price, USD/EUR exchange rate series obtained from Federal Reserve Economic Data database are used. They are daily data from 1.1999 to 11.2018. The original data series are converted into the corresponding logarithmic returns series. The descriptive statistics of original series as well as their log-returns is shown in Table 1.

	Original series		Log-returns series	
	EURUSD	Gold	EURUSD	Gold
Mean	8.4595×10^{-1}	8.7698×10^{2}	4.1499×10^{-6}	2.9990×10^{-4}
Median	8.1473×10^{-1}	9.1470×10^{2}	7.2823×10^{-5}	2.8825×10^{-4}
Mode	7.5506×10^{-1}	1.2375×10^{3}	0	0
Maximum	1.2092	1891	3.0031×10^{-2}	9.6416×10^{-2}
Minimum	6.2461×10^{-1}	2.5290×10^{2}	-4.6208×10^{-2}	-8.9128×10^{-2}
Std. deviation	1.2880×10^{-1}	4.7248×10^{2}	6.2578×10^{-3}	1.1370×10^{-2}
Skewness	8.4777×10^{-1}	1.0080×10^{-1}	-1.1956×10^{-1}	-5.8504×10^{-2}
Kurtosis	3.0101	1.6251	5.2323	11.0250
Numb of Obs	4865	4865	4864	4864

Table 1: Descriptive statistics of time series

First, we use the two return series to construct six portfolios in such a way that the share of their initial returns consisted of the returns of gold is 0, 0.2, 0.4, 0.6, 0.8 and 1 respectively. After generating corresponding of returns of those portfolio, we use these returns series to estimate the parameters of NIG a t-distributions using maximum likelihood estimation method, see [9]. The estimation procedures are performed in Matlab. Using the estimated values of distributions for returns of these portfolios, we calculate CVaR and VaR for confidence levels 97.5% and 99% respectively. For normal and Student t-distributions, as their inverse CDF exists, CVaR values are calculated according to 8. For NIG distribution, as its CDF does not exist, CVaR values are calculated as follows. First, the corresponding CDF is calculated by numerical integration. Then CVaR according to according to (8) is also calculated by numerical integration. Finally, CVaR of empirical distribution is the average all $r \leq \Phi^{-1}(1-\alpha)$. The results are reported in Table 2.

	CVaR975	VaR99	CVaR975	VaR99	CVaR975	VaR99	CVaR975	VaR99
w of g	CVaR norm	VaRnorm	CVaR t	VaR t	CVaR nig	VaR nig	CVaR HS	VaR HS
0	-0,0146	-0,0145	-0,0169	-0,0163	-0,0167	-0,0164	-0,0166	-0,0154
0,2	-0,0117	-0,0116	-0,0133	-0,0129	-0,0129	-0,0133	-0,0133	-0,0127
0,4	-0,0121	-0,0120	-0,0141	-0,0135	-0,0141	-0,0141	-0,0144	-0,0133
0,6	-0,0155	-0,0155	-0,0191	-0,0178	-0,0186	-0,0204	-0,0196	-0,0190
0,8	-0,0205	-0,0204	-0,0258	-0,0239	-0,0249	-0,0273	-0,0263	-0,0251
1	-0,0262	-0,0261	-0,0329	-0,0304	-0,0317	-0,0344	-0,0335	-0,0314

Table 2: Comparison of $CVaR_{97.5}$ and VaR_{99}

The results in Table 2 show that a portfolio of gold only is the most risky investment. It is possible to diversify the risk by creating portfolios of gold and eurodollar deposits. The optimum ratio of gold and eurodollar deposits is about 20% in gold and 80% in eurodollar deposits because both VaR and CVaR are (in absolute value) the lowest, leading to the smallest need for economic capital. This is valid for all types of distributions considered. If we compare the results for individual distributions for this ratio (0.2), the smallest need for economic capital is required when normal distribution is use d as an approximator for the distribution of returns. However, this inference needs to be taken with great caution as the normal distribution is not a good approximator for the (semi)fat tail property. Among the remaining three distributions, the smallest VaR in absolute value gives the empirical distribution, the largest VaR is for the NIG distribution. In contrast to the results of VaR, CVaR is the lowest one with NIG distribution. Otherwise, the values are the same for t-distribution and empirical distribution. In general, CVaR is greater than VaR for all these three distributions, resulting from the existence of heavy tails of returns (of course, VaR_{99} and $CVaR_{975}$ are the same when using normal distribution).

w of g	VaRs VaR norm	sVaR Norm	VaRs VaR t	sVaR t	VaRs VaR nig	sVaR nig	VaRs VaR HS	sVaR HS
0	-0,0145	-0,0145	-0,0163	-0,0163	-0,0164	-0,0164	-0,0154	-0,0154
0,2	-0,0116	-0,0168	-0,0129	-0,0191	-0,0133	-0,0197	-0,0127	-0,0200
0,4	-0,0120	-0,0191	-0,0135	-0,0219	-0,0141	-0,0227	-0,0133	-0,0234
0,6	-0,0155	-0,0215	-0,0178	-0,0248	-0,0204	-0,0257	-0,0190	-0,0267
0,8	-0,0204	-0,0238	-0,0239	-0,0276	-0,0273	-0,0287	-0,0251	-0,0301
1	-0,0261	-0,0261	-0,0304	-0,0304	-0,0344	-0,0344	-0,0314	-0,0314

Table 3: Subaditivity of VaR

Next, since VaR is generally not sub-additive, the sub-additivity of VaR is investigated. We calculate VaR_{99} for each portfolio for all type of distributions. These values are then compared with weighted sum of VaR_{99} of each component of portfolio. The results are displayed in Table 3. In the table, the VaR of portfolios are denoted as VaRs (VaR of sum of two assets), while the weighted sum of VaR are denoted as sVaR (sum of VaR of each components). The results we obtain show that in cases we analyze the sub-additivity property is maintained. The VaR of the entire portfolio is always smaller (in absolute value) than the weighted sum of individual VaRs.

w of g	CVaR norm	VaR norm	CVaR t	VaR t	CVaR nig	VaR nig	CVaR HS	VaR HS
0	-0,0161	-0,0160	-0,0193	-0,0185	-0,0154	-0,0156	-0,0197	-0,0142
0,2	-0,0137	-0,0136	-0,0166	-0,0158	-0,0134	-0,0135	-0,0176	-0,0122
0,4	-0,0138	-0,0138	-0,0170	-0,0159	-0,0151	-0,0148	-0,0371	-0,0152
0,6	-0,0165	-0,0164	-0,0203	-0,0183	-0,0198	-0,0194	-0,0569	-0,0196
0,8	-0,0206	-0,0205	-0,0274	-0,0233	-0,0265	-0,0261	-0,0766	-0,0264
1	-0,0254	-0,0253	-0,0340	-0,0287	-0,0342	-0,0329	-0,0964	-0,0334

Table 4: Var and CVaR of no correlation period 1999–2000

w of g	CVaR norm	VaR norm	CVaR t	VaR t	CVaR nig	VaR nig	CVaR HS	VaR HS
0	-0,0189	-0,0189	-0,0243	-0,0226	-0,0231	-0,0227	-0,0300	-0,0214
0,2	-0,0147	-0,0147	-0,0180	-0,0170	-0,0179	-0,0176	-0,0242	-0,0181
0,4	-0,0166	-0,0165	-0,0204	-0,0192	-0,0187	-0,0183	-0,0312	-0,0200
0,6	-0,0232	-0,0231	-0,0289	-0,0271	-0,0259	-0,0254	-0,0526	-0,0249
0,8	-0,0317	-0,0316	-0,0393	-0,0369	-0,0351	-0,0344	-0,0740	-0,0326
1	-0,0410	-0,0408	-0,0507	-0,04768	-0,04511	-0,0442	-0,0955	-0,0425

Table 5: VaR and CVaR of high negative correlation period 2007–2009

Finally, we investigate the possible effect of correlation relationship between individual components on the optimal structure of portfolios. We choose two periods 1999-2000 and 2007-2009. The first period was characterized by an almost zero average correlation of USD and gold, whereas in the second period the average correlation was significantly negative (see Figure 2). First we use data from these two periods to reestimate parameters of each considered distribution. Then, VaR_{99} and $CVaR_{97.5}$ were calculated for both periods with the same methodology described above. The results are shown in Tables 4 and 5. Due to the reduction of data for parameter estimation, the estimated parameter of distributions are slightly changed. These changes lead to the increase of calculated VaR and CVaR. The biggest change is observed in the case of historical simulation for obtaining the empirical distribution. The results may raise the question of the suitability of historical simulation method for calculating VaR and CVaR when it has to work with lesser quantity of historical data. Regarding the main focus of our investigation in this part, the results in these two tables show that the (average) strength of gold-USD return correlation does not have a significant impact on the optimal structure of portfolio measured by VaR and CVaR. The optimum ratio eurodollar deposits and gold is again around 0.8:0.2. Even for higher negative correlation (Table 5), VaR and CVaR are always larger (in absolute value) than in the case of almost zero correlation between the two assets.

6 Conclusion

We examine the possibility to use VaR and CVaR as an instrument for portfolio optimization. For this purpose, we construct several portfolios consisted of eurodollar deposits and gold. The returns of these portfolios are modeled by normal, Student t, NIG and empirical distributions. It turns out that for both VaR and CVaR, as a measure of risk, the optimal ratio of gold to dollar deposits is about 0.2. Because tails of Student t, NIG and empirical distributions are thicker than those of normal distribution, CVaR is lower than VaR, so more economic capital is needed. When investing the role of correlation among assets in a portfolio, we have found that if the correlation is close to zero, there is no significant effect on the left end of the distribution. Negative correlation requires more economic capital because VaR and CVaR are lower than in the (average) zero correlation. This is a somewhat surprising result that may be explained by uneven drops and rises (because correlation is only the measure of linear dependencies), or numerical errors. However, this situation will require further investigation. Although VaR is not generally sub-additive, this property is maintained in all our cases. This fact supports the claim that the VaR sub-additivity is not impaired in known practical situations.

Acknowledgements

Author Jiří Málek acknowledges the financial support of Czech Science Foundation with grant GAČR 18-05244S "Innovative Approaches to Credit Risk Management" and Institutional support IP 100040/1020. Author Tran van Quang is grateful for the financial support of grant GAČR 18-05244S "Innovative Approaches to Credit Risk Management" of Czech Science Foundation and Institutional support IP 100040/1060.

References

[1] Barndorff-Nielsen, O. E., Blæsild, P., Jensen, J. L., Sørensen, M. (1985). The fascination of sand. In:*A Celebration of Statistics* , (Atkinson, A. C., Fienberg, S. E.,ed.), Springer Verlag, New York, pp. 57—87.

[2] Barndorff-Nielsen, O. E.(1998). Processes of normal inverse Gaussian type, *Finance and Stochastics* 2, 41–68.

[3] Bibby, B. M., Sorensen, M. (2003). Hyperbolic processes in finance. In:*Handbook of Heavy Tailed Distributions in Finance*, (S. Rachev, ed.), Elsevier Science Amsterdam, pp. 211–248.

[4] Borak, S., Misiorek, A., and Weron, R. (2010). Models for Heavy-tailed Asset Returns, *MPRA paper No. 25494*, September 2010.

[5] Chu-Hsiung Lin, Shan-Shan Shen (2006). Can the student t-distribution provide accurate value at risk? *The Journal of Risk Finance*, 7(3), pp.292-300.

[6] Eberlein, E. and Hammerstein, E. A. (2003). Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting Cases and Approximation of Processes. Working paper University of Freiburg, 2003.

[7] Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution.*Statistics & Probability Letters 57* (2002), pp. 43–52.

[8] Kumari, S. N. and Tan, A. (2013). Characterization of Student's t-distribution with some Application to Finance. *Mathematical Theory and Modeling*, 3(10), 2013.

[9] Málek, J. and Tran Van Quang (2016). Evaluating the ability to model the heavy tails of asset returns of two alternative distributions, *Mathematical Methods in Economics, Proceedings of the 34th International Conference, Liberec, Czech Republic*, September 6th – 9th 2016, pp. 528-533.

[10] Misiorek, A. and Weron, R (2010). Heavy-tailed distributions in VaR calculations, *HSC Research Report Series 2010*.

[11] Prause, K. (1999). *The generalized hyperbolic model: Estimation, financial derivatives and risk measures.* University of Freiburg, Doctoral Thesis.

Optimal depot-dump routes design for municipal waste collection vehicles

Šárka Michalcová[1], Petr Kozel[2], Lucie Orlíková[3]

Abstract. The design of vehicle round routes for the transport network maintenance often reduces to solving that situations, when the individual routes have the same starting and ending point. This is a type of depot-depot route. However, the starting and ending point are usually different in practice. This is a situation, in which the waste collection vehicles leave their depot, collect waste gradually and then direct to the dump to unload the cargo. That is a case of depot-dump round route. The presented paper is devoted to the introduction of automated input data editing, which allow to solve this task using mathematical model ATSP (Asymmetric Traveling Salesman Problem). This contribution presents both sample solutions on trivial data and solutions of real situations of particular commercial company providing municipal waste collection in Třinec.

Keywords: linear programming, vehicle routing problem, Hamiltonian circuit problem, asymmetric traveling salesman problem (ATSP), municipal waste collection, ArcGIS.

JEL classification: C61
AMS classification: 90C05

1 Introduction

Maintenance and operation management of the transport infrastructure are the main activities for logistic companies, technical services and other subjects that ensure the customer service within the transport network. With respect to efficient and economical performance, the main target is to design the service vehicle route so that the chosen optimization criterion (mostly the total distance traveled in kilometers) is considered while respecting all significant limitations of real-time traffic.

In practice, there is a whole range of these real limitations that may include some transport network particularities (the existence of mixed transport network) [3], requirements regarding not exceeding the capacity of service vehicles (decomposition) [7], [4] etc. One of the real requirements is to design a route for a service vehicle which starting and ending point differ. First of all, this requirement is often used for planning routes of municipal waste collection vehicles where starting point is the depot of service vehicles and targeting point is a place of the dump. Naturally, these both points are not identical.

Assuming that the above-described situation is modeled using the task of vertex service, the service route design can be based on determining of the minimum Hamiltonian circuit (Traveling Salesman Problem TSP, Asymmetric Traveling Salesman Problem ATSP respectively) with the specific modification of input data. This approach is described in detail in the following text of this contribution.

2 Formulation of problem

As already mentioned, the theoretical concept for the depot-dump route design is based on the task of finding the minimum Hamiltonian circuit, which is type of the depot-depot. This is round, where both starting and end points coincide and the formulation of this problem is as follows.

2.1 Formulation of the depot-depot problem

Let the mixed graph be specified by $N(V, E, l)$ with three attributes. The first one $V = 1, ..., m$ denotes a set of vertices, $E = 1, ..., n$ is a set of edges (directed and undirected edges) and the last one l_{ij} represents an edge evaluation ij (the road length in kilometers), $i, j \in V$. Example of the graph N is shown in Figure 1 and it will be

[1]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, sarka.michalcova@vsb.cz
[2]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, petr.kozel@vsb.cz
[3]Technical University of Ostrava/Faculty of Mining and Geology, Institute of Geoinformatics, 17. listopadu 15, Ostrava - Poruba, Czech republic, lucie.orlikova@vsb.cz

used for the sample task. The goal is to find the route with the minimal length and which passes through all edges of the graph just once except of the initial vertex, which is the end vertex simultaneously. To solve this problem, one advisable approach is to determine a minimum Hamiltonian cycle.

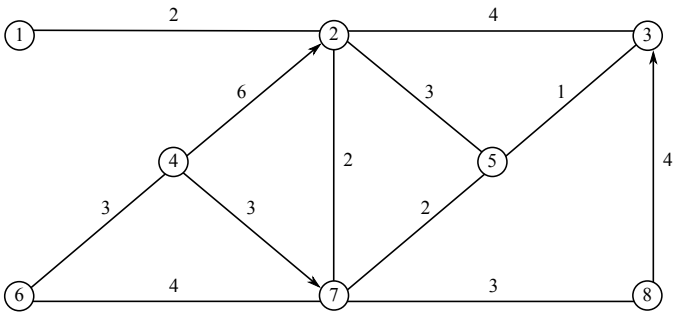


Figure 1: Diagram showing the migraph N

In order to find a minimum Hamiltonian cycle, it is necessary that the given graph must be Hamiltonian [6]. Nonetheless, graphs corresponding to the real networks often do not fulfill this condition and must be transformed to the complete graph. The matrix of distances c_{ij} between individual vertices of the complete graph can be calculated for instance by employing the *Floyd-Warshall algorithm*. Both the matrix of edge evaluation l_{ij} and the matrix of distances c_{ij} corresponding to the migraph N in Figure 1 are listed in Table 1.

Table 1: Matrix of edge evaluation and distance matrix in the migraph N

(a) The matrix of edge evaluation l_{ij}

l_{ij}	1	2	3	4	5	6	7	8
1	0	2	-	-	-	-	-	-
2	2	0	4	-	3	-	2	-
3	-	4	0	-	1	-	-	-
4	-	6	-	0	-	3	3	-
5	-	3	1	-	0	-	2	-
6	-	-	-	3	-	0	4	-
7	-	2	-	-	2	4	0	3
8	-	-	4	-	-	-	3	0

(b) The distance matrix c_{ij}

c_{ij}	1	2	3	4	5	6	7	8
1	0	2	6	11	5	8	4	7
2	2	0	4	9	3	6	2	5
3	6	4	0	10	1	7	3	6
4	7	5	6	0	5	3	3	6
5	5	3	1	9	0	6	2	5
6	8	6	7	3	6	0	4	7
7	4	2	3	7	2	4	0	3
8	7	5	4	10	5	7	3	0

Using the distance matrix c_{ij}, it is already possible to find a minimum Hamiltonian cycle and the following form of linear mathematical model can be used for small dimensional tasks [2].

$$\text{Min} \sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} x_{ij} \cdot c_{ij} \tag{1}$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{m} x_{ij} = 1, \text{ for } j = 1, ..., m \tag{2}$$

$$\sum_{\substack{k=1 \\ j \neq k}}^{m} x_{jk} = 1, \text{ for } j = 1, ..., m \tag{3}$$

$$y_j - y_i + m \cdot x_{ij} \leq m - 1, \text{ for } i = 2, ..., m, \ j = 2, ..., m, \ i \neq j \tag{4}$$

$$x_{ij} \in \{0, 1\}, \text{ for } i = 1, ..., m, \ j = 1, ..., m, \ i \neq j \tag{5}$$

$$y_i \in \mathbb{Z}_0^+, \text{ for } i = 2, ..., m \tag{6}$$

The bivalent variable x_{ij} represents the edge inclusion (resp. non-inclusion) ij to the route. The objective function (1) denotes the total distance of the route. Conditions (2) and (3) ensure that just one edge enters each vertex and just one edge exits each vertex, too. Conditions (4) do not allow creation of subcycle. Conditions (5) and (6) determine the domain of variables x_{ij} and y_i. Mathematical model (1) – (6) represents an exact way for finding a minimum Hamiltonian circuit. If we apply this approach to the sample example of the network N, we get the solution corresponding to the sequence of vertices 1-6-4-7-8-3-5-2-1, resp. 1-2-7-6-4-7-8-3-5-2-1 with the length **27** km.

Just introduced approach can be used in that situation when both starting and ending points of the route coincide. Furthermore, we proceed to formulate a problem when the starting and ending points of the route are different.

2.2 Formulation of the depot-dump problem

Let the mixed graph be specified by $N(V, E, l)$ with three attributes. The first one $V = 1, ..., m$ denotes a set of vertices, $E = 1, ..., n$ is a set of edges (directed and undirected edges) and the last one l_{ij} represents an edge evaluation ij (the road length in kilometres), $i, j \in V$. At the vertex 1, there is a depot, which the service vehicle leaves. At the vertex m, there is a dump where the service vehicle is heading. The goal is to determine such route with the minimal length, which passes through all vertices of the graph just once except the initial vertex where the route begins and the vertex m where the route ends.

3 Input data modification

To solve just described problem, the above mentioned mathematical model (1) – (6) can be employed after modification of input data. This will be described in detail in the following.

One fictitious vertex f is included into the set of vertices $V = 1, ..., m$ and $m + 1$ fictitious edges are included into the set of edges $E = 1, ..., n$, too. As a result of this modification, both the new vertex set $\overline{V} = \{1, ..., m\} \cup \{f\}$ and the new edge set $\overline{E} = \{1, ..., n\} \cup \{(1, f), (f, 1), \{2, f\}, ..., \{m, f\}\}$ can be described. These auxiliary sets of edges and vertices define a new migraph $\overline{N}$. This situation is illustrated at the sample example and shown in Figure 2. The fictitious vertex f is plotted twice and the fictitious edges are represented by the dashed line.

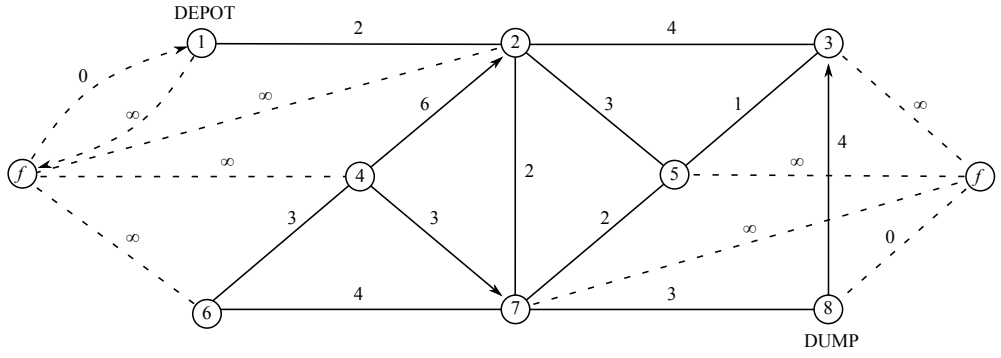


Figure 2: Diagram showing the migraph $\overline{N}$

Both the matrix of the edge evaluation $\overline{l_{ij}}$ and the distance matrix $\overline{c_{ij}}$ corresponding to the migraph $\overline{N}$ of Figure 2 are shown in Table 2.

In general, modification of the distance matrix c_{ij} corresponding to the migraph N on the matrix of distances $\overline{c_{ij}}$ corresponding to the migraph $\overline{N}$ can be written as follows:

$$\overline{c_{ij}} = \begin{cases} c_{ij} & \text{for} \quad i = 1, ..., m\,, j = 1, ..., m\,, \\ \infty & \text{for} \quad i = 1, ..., m-1\,, j = f\,, \\ \infty & \text{for} \quad i = f\,, j = 2, ..., m-1\,, \\ 0 & \text{for} \quad i = f\,, j = 1\,, \\ 0 & \text{for} \quad i = f\,, j = m\,, \\ 0 & \text{for} \quad i = m\,, j = f\,. \end{cases}$$

Table 2: Matrix of edge evaluation and distance matrix in the migraph $\overline{N}$

(a) The matrix of edge evaluation $\overline{l_{ij}}$

$\overline{l_{ij}}$	1	2	3	4	5	6	7	8	f
1	0	2	-	-	-	-	-	-	∞
2	2	0	4	-	3	-	2	-	∞
3	-	4	0	-	1	-	-	-	∞
4	-	6	-	0	-	3	3	-	∞
5	-	3	1	-	0	-	2	-	∞
6	-	-	-	3	-	0	4	-	∞
7	-	2	-	-	2	4	0	3	∞
8	-	-	4	-	-	-	3	0	0
f	0	∞	∞	∞	∞	∞	∞	0	0

(b) The distance matrix $\overline{c_{ij}}$

$\overline{c_{ij}}$	1	2	3	4	5	6	7	8	f
1	0	2	6	11	5	8	4	7	∞
2	2	0	4	9	3	6	2	5	∞
3	6	4	0	10	1	7	3	6	∞
4	7	5	6	0	5	3	3	6	∞
5	5	3	1	9	0	6	2	5	∞
6	8	6	7	3	6	0	4	7	∞
7	4	2	3	7	2	4	0	3	∞
8	7	5	4	10	5	7	3	0	0
f	0	∞	∞	∞	∞	∞	∞	0	0

Furthermore, using the mathematical model (1) – (6), a minimum Hamiltonian circuit can be found in the migraph $\overline{N}$ and the result of this task is a circuit ride 1-i- ... -i-m-f. If we apply just presented approach to the sample example, we get solution corresponding to the sequence of vertices 1-2-3-5-7-6-4-8-f-1, resp. 1-2-3-5-7-6-4-8-f-1 with the length **22** km. The vertex 8 in which the dump is located, is at the end of the route. In addition, the only fictitious vertex f (which does not exists in real) is located beyond the vertex 8 and also the vertex 1 as a result of the path from the fictitious vertex f to the vertex 1. It corresponds to ending the round route, but this route (8-f-1) has length of **0** km.

4 Numerical experiments

In order to test the usability of the proposed approach computational experiments were conducted. In the first phase of the research, the free and open-source database TSPLIB 95 was used for testing ATSP (Asymmetric Traveling Salesman Problem) [8]. In total twelve task, which are exactly solvable using the mathematical model (1) – (6), were selected from this database. As follows, in the second phase this approach was tested using input data from the municipal waste in Písečná, the Moravian-Silesian Region.

4.1 TSPLIB 95 database

Table 3 summarizes results of numerical experiments on 12 selected tasks from the TSPLIB 95 database. Each task includes its designation, number of cities, the length of the depot-depot round route (in km) and also the length of the depot-dump round route (in km).

Table 3: Results of numerical experiments on TSPLIB 95 database tasks

(a) Tasks br17 – ftv38

Code	Cities	Depot-Depot [km]	Depot-Dump [km]
br17	17	39	34
ft53	53	6 905	6 196
ft70	70	38 673	37 655
ftv33	34	1 286	1 275
ftv35	36	1 473	1 443
ftv38	39	1 530	1 511

(b) Tasks ftv44 – ry48p

Code	Cities	Depot-Depot [km]	Depot-Dump [km]
ftv44	45	1 613	1 614
ftv47	48	1 776	1 754
ftv55	56	1 608	1 613
ftv64	65	1 839	1 840
ftv70	71	1 950	1 958
ry48p	48	14 442	14 070

For all computational experiments, the sequence of individual vertices in the resulting route was checked. As required by the aforementioned conditions, the vertex at which the dump is located was the last one in all cases. At the same time, results of all experiments are as follows:

$$\sum_{\substack{i\in V\\ i\neq j}}\sum_{j\in V} c_{ij}\cdot e_{ij} \leq \sum_{\substack{i\in \overline{V}\\ i\neq j}}\sum_{j\in \overline{V}} \overline{c_{ij}}\cdot \overline{e_{ij}} + \overline{c_{m1}}. \tag{7}$$

The quantity e_{ij} represents the depot-depot edges and $\overline{e_{ij}}$ denotes the depot-dump edges.

4.2 Real problem from practice

As mentioned above, the proposed approach was verified using input data from the municipal waste collection environment in Písečná in the Moravian-Silesian Region. There are 262 land registry numbers on the territory of municipality, where a container for municipal waste collection is registered. This set of 262 customers was decomposed into 5 subsets corresponding to 5 service vehicle routes with respect to the given average filling of containers. After that, it is necessary to plan the depot-dump service vehicle routes for these subsets assuming that both the location, which service vehicles leave (depot), and the location, where vehicles collect waste (dump), are exactly specified. The mentioned decomposition was realized using the Sweep Algorithm [7]. Geocoding of individual customers and calculations of distance matrices were performed using ArcGIS [1], [5]. The results of numerical experiments with real data are presented in Table 4.

Table 4: Results of numerical experiments with real data

Service route	The number of customers	Depot-Depot [km]	Depot-Dump [km]
1	52	39.213	25.195
2	52	39.628	25.229
3	52	34.936	21.329
4	52	40.866	27.143
5	54	37.963	23.827

As well as described experiments using TSPLIB database, the vertex, at which is the dump located, is the last vertex of the route in all cases. Taking to account the fact that the shortest route length for the dump-depot case is **14.485** km, it can be verified that the relationship (7) is valid for all 5 routes.

5 Conclusion

The presented paper was devoted to the introduction of automatic modification of input data, which enable to solve the task of depot-dump type using the mathematical model for determining the minimum Hamiltonian circuit. Chapter 2 was in detail dedicated to problem formulation and specific data modification was described in Chapter 3. Computational experiments with the proposed approach were presented in Chapter 4 both on test tasks using TSPLIB 95 database and on tasks with real input data from municipal waste collection environment. All computational experiments were realized in the optimization solver for linear programming Xpress-IVE and software ArcGIS. To sum up, presented input data modification can be used to solve the task of the depot-dump route designing.

Acknowledgements

This paper was supported within Op. Prog. Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296 and programme for the support of applied research and experimental development, Technology Agency of the Czech Republic (EPSILON): 'Effective approaches to economical and adaptable systems of maintenance and operation of transport networks' TH02010930.

References

[1] Allen, D., W. (2013). *GIS Tutorial 2: Spatial Analysis Workbook*. New York: Esri Press.

[2] Janáček, J. (2003). *Matematické programování*. Žilina: Žilinská univerzita.

[3] Kozel, P., Michalcová, Š. (2015). The use of linear programming in solving service tasks of mixed transport networks. In: *Proceedings of 11th International Conference Strategic Management and its Support by Information Systems 2015* (pp. 274–279). Uherské Hradiště.

[4] Kozel, P., Orlíková, L., Pomp, M., Michalcová, Š. (2018). Application of the p-Median Approach for a Basic Decomposition of a Set of Vertices to Service Vehicles Routing Design. In: *Proceedings of 36th International Conference Mathematical Methods in Economics 2018* (pp. 252–257). Jindřichův Hradec.

[5] Ormsby, T., Napoleon, E. et al. (2010). *Getting to Know ArcGIS Desktop*. New York: Esri Press.

[6] Palúch, S. (2008). *Algoritmická teória grafov*. Žilina: Žilinská univerzita.

[7] Pomp, M., Kozel, P., Michalcová, Š., Orlíková, L. (2017). Using the Sweep Algorithm for decomposing a set

of vertices and subsequent solution of the traveling salesman problem in decomposed subsets. In: *Proceedings of 35th International Conference Mathematical Methods in Economics 2017* (pp. 584–589). Uherské Hradiště.

[8] Reinelt, G. (2013). *TSPLIB 95*. Universität Heidelberg. `https://wwwproxy.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/`.

[9] Rybička, J. (2003). *LaTeXpro začátečníky*. Brno: Konvoj.

[10] *XPRESS-Mosel "User guide"*. (2005). United Kingdom: Dash Associates.

A Nonparametric Bootstrap Comparison of Variability of Robust Regression Estimators

Jan Kalina[1], Nicole Tobišková[2], Jan Tichavský[3]

Abstract. While various robust regression estimators are available for the standard linear regression model, performance comparisons of individual robust estimators over real or simulated datasets seem to be still lacking. In general, a reliable robust estimator of regression parameters should be consistent and at the same time should have a relatively small variability, i.e. the variances of individual regression parameters should be small. The aim of this paper is to compare the variability of S-estimators, MM-estimators, least trimmed squares, and least weighted squares estimators. While they all are consistent under general assumptions, the asymptotic covariance matrix of the least weighted squares remains infeasible, because the only available formula for its computation depends on the unknown random errors. Thus, we take resort to a nonparametric bootstrap comparison of variability of different robust regression estimators. It turns out that the best results are obtained either with MM-estimators, or with the least weighted squares with suitable weights; the latter estimator is especially recommendable for small sample sizes.

Keywords: robustness, linear regression, outliers, bootstrap, least weighted squares.

JEL classification: C14
AMS classification: 62G35

1 Introduction

In the standard linear regression model, the least squares estimator is notoriously known to suffer from the presence of outlying values (outliers) in the data. Therefore, numerous robust regression estimators have been proposed as resistant alternatives [6]. In general, it is expected from a reliable regression estimator that it is consistent and has a low variability. The latter means that the variances of individual regression parameters are small, or (more precisely) that the covariance matrix of the asymptotically normal distribution remains small in a certain (selected) sense, e.g. has a small determinant. The variance of regression estimates is an important characteristic and there have been successful attempts to interpret it in particular economic tasks [5]. Therefore, it remains important to compare the variability of different robust regression estimators.

An explicit form of the asymptotic covariance matrix is available for least trimmed squares, S-estimators, or MM-estimators. All these estimators are known as highly robust, because they (may potentially) achieve a high breakdown point, i.e. a high resistance against severe outliers in the data. We are primarily interested in the least weighted squares estimator, which turns out to possess several appealing properties. However, there seem no comparisons of the covariance matrix of the least weighted squares estimator with that of other highly robust estimators. The asymptotic covariance matrix for the LWS derived by [14] depends on the unknown regression errors and its estimate seems not to be available. A thorough comparison of the performance of the LWS with different weight function seems also missing.

Bootstrap estimation (bootstrapping, resampling with replacement) has become a popular tool (not only) in the task to estimate the variance (or covariance matrix) of estimators in different contexts. Practical approaches to bootstrapping were developed and investigated by [3]. Incorporating the basic principles of bootstrapping, one may develop a great variety of resampling techniques that provide us with new possibilities of analyzing data by means of residual bootstrap, semiparametric bootstrap, Bayesian bootstrap etc.

The aim of the paper is to compare the covariance matrix of several robust regression estimators. Section 2 recalls several prominent robust regression estimators and lists 9 their versions, which are used in our computations. These include also MM-estimators [16], which are currently perceived with an unconcealed optimism [4] in the regression context, because they are theoretically known to combine high robustness with high efficiency; they

[1]The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Praha 8, Czech Republic, kalina@cs.cas.cz

[2]The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Praha 8, Czech Republic, nicole.tobiskova@gmail.com

[3]The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Praha 8, Czech Republic, tichavsk@seznam.cz

were actually designed in this way with the aim to overcome the low efficiency of S-estimators. Keeping in mind the increasing importance of the bootstrap (resampling) methodology in econometrics [10], we perform an extensive numerical bootstrap study comparing the covariance matrix obtained by highly robust estimators in Section 3. Finally, Section 4 concludes the paper.

2 Robust regression

We consider the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \ldots, n, \tag{1}$$

where the continuous response $Y_1, \ldots, Y_n$ is explained by the total number of p regressors. The errors (disturbances) $e_1, \ldots, e_n$ are assumed to be independent and identically distributed, but not necessarily normally distributed. We are strongly interested in the least weighted squares estimator [15], which is far from being well known in the econometric community, and therefore we recall its definition in Section 2.1. Section 2.2 gives a list of 9 estimators, which will be used in the numerical study.

2.1 Implicitly weighted estimators

The least trimmed squares (LTS) estimator of β (see e.g. [12]) is nowadays probably the most common highly robust estimator in linear regression. The user must select the value of a trimming constant h ($n/2 \leq h < n$). We will need to denote residuals corresponding to a particular $b = (b_0, \ldots, b_p)^T \in \mathbb{R}^{p+1}$ as

$$u_i(b) = Y_i - b_0 - b_1 X_{i1} - \cdots - b_p X_{ip} \tag{2}$$

and order statistics of their squares as

$$u^2_{(1)}(b) \leq \cdots \leq u^2_{(n)}(b). \tag{3}$$

The estimator denoted as LTS-α is obtained as

$$\operatorname*{arg\,min}_{b \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^{h} u^2_{(i)}(b), \tag{4}$$

where $\alpha = h/n \in [1/2, 1)$. The LTS estimator may attain a high robustness but cannot achieve a high efficiency. We may consider the LTS as an implicitly weighted estimator, namely as a special case of the least weighted squares with weights equal only to 0 or 1.

The least weighted squares (LWS) estimator (see e.g. [15]) for the model (1) is motivated by the idea to down-weight potential outliers, but remains much less known compared to the LTS. The estimator may achieve a high breakdown point (with properly selected weights) and is robust to heteroscedasticity [15]. The idea of the LWS, i.e. the implicit weighting based on ranks of residuals, is successful in a variety of recent applications e.g. in tourism management [9] or forensic anthropology [7]. An extension of the LWS to nonlinear regression was presented in [8]. However, we are not aware of any comparison of the performance of the LWS with that of MM-estimators.

The definition of the LWS exploits the concept of weight function, which is defined as a function $\psi : [0, 1] \to [0, 1]$; it must be non-increasing and continuous on $[0, 1]$ with $\psi(0) = 1$ and $\psi(1) = 0$. The weight function is assumed to have both one-sided derivatives existing in all points of $(0, 1)$, where the one-sided derivatives are bounded by a common constant; also, the existence of a finite left derivative in 0 and finite right derivative in point 1 is assumed [15]. The LWS estimator with a given ψ is defined as

$$\mathbf{b}^{LWS} = (b_0^{LWS}, b_1^{LWS}, \ldots, b_p^{LWS})^T = \operatorname*{arg\,min}_{b \in \mathbb{R}^{p+1}} \sum_{k=1}^{n} \psi\left(\frac{k-1}{n}\right) u^2_{(i)}(b). \tag{5}$$

Some possible choices of weights for the LWS will be listed in Section 2.2. For the computation of the LWS estimator, we propose an algorithm presented as Algorithm 1, which is denoted as FAST-LWS. This algorithm, not implemented in any publicly available software yet, was obtained by adapting the FAST-LTS algorithm of [12]. The algorithm exploits the weights least squares (WLS), i.e. a weighted version of the least squares with fixed weights; this (very standard) estimator should not be confused with the LWS. The explicit formula for the WLS estimator with fixed weights $w_1, \ldots, w_n$ is $(X^T W X)^{-1} X^T W Y$, where W is a diagonal matrix $W = \mathsf{diag}\{w_1, \ldots, w_n\}$.

Algorithm 1 FAST-LWS estimator of β in the linear regression model (1)

Input: $X_1, \ldots, X_n$, where $X_i \in \mathbb{R}^p$ for each $i = 1, \ldots, n$
Input: $Y_1, \ldots, Y_n$, where $Y_i \in \mathbb{R}$ for each $i = 1, \ldots, n$
Input: Weight function ψ
Input: $J > 0$
Input: $\varepsilon > 0$
Input: Magnitudes of weights $w^1 \geq \cdots \geq w^n$
Output: Optimal permutation of $w^1, \ldots, w^n$ denoted as $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^n)^T$
Output: b_{LWS}

for $j = 1$ to J **do**
  $m_{j0} := +\infty$
  Select randomly $p + 1$ observations
  Estimate β by $b_{j0} = (b^0_{j0}, b^1_{j0}, \ldots, b^p_{j0})^T$ obtained as the least squares estimator in (1) using exactly the selected $p + 1$ observations
  $w_{j,0} := (1, \ldots, 1)^T \in \mathbb{R}^n$
  $k := 0$
  repeat
    $u_i(b_{jk}) := Y_i - b^0_{jk} - b^1_{jk} X_{i1} - \cdots b^p_{jk} X_{ip}$ for $i = 1, \ldots, n$
    $o_1, \ldots, o_n :=$ vector of order statistics computed from $u_1(b_{jk}), \ldots, u_n(b_{jk})$
    $w_{j,k+1} := (w^1_{j,k+1}, \ldots, w^n_{j,k+1})^T$ with

$$w^i_{j,k+1} := \psi\left(\frac{o_i - 1}{n}\right) \quad \text{for } i = 1, \ldots, n \tag{6}$$

    Estimate β by $b_{j,k+1} = (b^0_{i,k+1}, b^1_{i,k+1}, \ldots, b^p_{i,k+1})^T$ obtained as the WLS estimator with weights $w_{j,k+1}$
    $m_{j,k+1} := \sum_{i=1}^n w^i_{j,k+1} u^2_{(i)}(b_{j,k+1})$
    $k := k + 1$
  until $m_{jk} > m_{j,k-1} + \varepsilon$
  $\tilde{b}_j = (\tilde{b}^1_j, \ldots, \tilde{b}^n_j)^T := b_{j,k-1}$
  $\tilde{w}_j = (\tilde{w}^1_j, \ldots, \tilde{w}^n_j)^T := w_{j,k-1}$
end for
$j^* := \arg\min_{j=1,\ldots,J} \sum_{i=1}^n \tilde{w}_j u^2_{(i)}(\tilde{b}_j)$
$\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^n)^T := (\tilde{w}^1_{j^*}, \ldots, \tilde{w}^n_{j^*})^T$
$b_{LWS} :=$ weighted least squares estimator in (1) with weights $\tilde{w}_1, \ldots, \tilde{w}_n$

2.2 Estimators used in the numerical study

All estimators used in this paper except for the LWS are implemented in the `robustbase` package [13] of R software [11]. For each estimator, the following list includes also the name of the function from this package, which we used for its computation. For all versions of the LWS, we use our own implementation of Algorithm 1 programmed in R software.

1. Least squares, computed by the function lm of R software.
2. S-estimator [2] with breakdown point equal to 0.5. In the computations, we use the function lmrob.S of [13].
3. MM-estimator with breakdown point equal to 0.5 and with efficiency equal to 0.95. In the computations, we use the function lmrob of [13].
4. LTS-0.55. In the computations, we use the function ltsReg of [13].
5. LTS-0.80.
6. LWS-A defined as LWS with linear weights, i.e. weights generated by the weight function

$$\psi(t) = 1 - t, \quad t \in [0, 1]. \tag{7}$$

7. LWS-B defined with weights generated by the logistic function

$$\psi(t) = \frac{1 + \exp\{-s/2\}}{1 + \exp\{s(t - \frac{1}{2})\}}, \quad t \in [0, 1], \tag{8}$$

for a given constant $s > 0$ responsible for the shape of the logistic curve, while we consider $s = 10$ in all the computations.

8. LWS-C defined with trimmed linear weights generated for a fixed $\tau \in [1/2, 1)$ by

$$\psi(t) = \left(1 - \frac{t}{\tau}\right) \cdot \mathbb{1}[t < \tau], \quad t \in [0, 1], \tag{9}$$

where $\mathbb{1}[.]$ denotes an indicator function. Here, τ is a parameter corresponding to trimming, i.e. there are $\lfloor \tau n \rfloor$ observations retained and the remaining observations are ignored, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$; this is analogous to α for the LTS. We use $\tau = 0.75$ in all the computations.

9. LWS-D defined with weights exploiting the so-called error function (erf), defined by the weight function

$$\psi(t) = 1 - \mathsf{erf}(t) = 1 - \frac{2}{\sqrt{\pi}} \int_0^t \exp\{-x^2\} dx; \tag{10}$$

we use the package `NORMT3` of R software to evaluate the error function.

3 A nonparametric bootstrap numerical study

A numerical study in R software [11] is performed with the aim to compare the variability of several robust regression estimators. We consider a simple setup in order for the (computationally intensive) bootstrap study to be feasible. We use with $p = 1$, $n = 30$ or $n = 60$, and $N = 100$ different datasets. For each dataset, the regressor is randomly generated from uniform distribution $\mathsf{U}(0, 10)$, i.e. over the interval $(0, 10)$. We always use $\beta = (2, 1)^T$ and $\sigma^2 = 1$. The disturbances are generated from normal distribution as $e_i \sim \mathsf{N}(0, \sigma^2)$ with $\sigma = 0.3$. We consider each of the N datasets itself, together with $R = 100$ contaminated versions according to Contamination I and $R = 100$ contaminated versions according to Contamination II. Each contamination considers c % of randomly selected observations (using $c = 0$ or $c = 10$ or $c = 30$) in the following ways:

- Contamination I. We randomly choose c % of observations. The regressor is generated for the randomly selected points from $\mathsf{U}(0.5, 2)$ and the corresponding response according to (1) with the error generated from $\mathsf{N}(0, 5)$.
- Contamination II. We randomly choose c % of observations. Then, a value m is generated from $\mathsf{U}(0, 10)$. For each of the selected observations, a specific value v is generated from $\mathsf{U}(-0.5, 0.5)$ and the regressor is replaced by $m + v$; the response is replaced by a value obtained according to (1) with the error generated from $\mathsf{U}(3, 6)$.

For each of the contaminated datasets, $S = 100$ bootstrap samples are created and evaluated according to Algorithm 2. This is nonparametric bootstrap formulated for the LWS estimator and it is used in the same way for each of the estimators of Section 2.2. For each bootstrap sample, each of 9 estimators presented in Section 2.2 are computed. The results are then averaged over the total number of S bootstrap samples. From the averaged results computed over the datasets in each situation, the covariance matrix is obtained. Table 1 presents the results of the numerical experiment for $n = 30$ and Table 2 for $n = 60$. There, we evaluate average values of $\mathsf{var}\, b_0$ and $\mathsf{var}\, b_1$, where $b = (b_0, b_1)^T$ corresponds to each of the estimators under consideration.

Table 1 Estimated variances of b_0 and b_1 obtained by the nonparametric bootstrap study for $n = 30$. Here, $b = (b_0, b_1)^T$ are estimates of $\beta = (\beta_0, \beta_1)^T$ by means of 9 different estimators of Section 2.2.

		No contam.		Contam. I 10 % of outliers		Contam. I 30 % of outliers		Contam. II 10 % of outliers		Contam. II 30 % of outliers	
		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	LS	**0.18**	**0.0030**	0.57	0.011	1.34	0.032	0.32	0.0063	0.57	0.024
2	S-est.	0.30	0.0073	0.28	0.010	0.59	0.013	0.23	0.0100	0.36	0.012
3	MM-est.	**0.18**	0.0034	**0.20**	0.006	0.83	0.014	0.22	0.0074	**0.28**	**0.012**
4	LTS-0.55	0.19	0.0061	0.23	0.008	0.49	0.012	0.20	**0.0051**	0.33	0.018
5	LTS-0.80	0.20	0.0033	0.21	**0.005**	0.69	0.013	0.20	0.0057	0.43	0.019
6	LWS-A	0.19	0.0039	0.30	0.007	0.69	0.014	**0.18**	0.0064	0.41	0.018
7	LWS-B	0.23	0.0055	0.31	0.009	**0.45**	**0.011**	0.21	0.0086	0.35	0.014
8	LWS-C	0.26	0.0076	0.30	0.010	0.47	0.013	0.28	0.0100	0.34	**0.012**
9	LWS-D	0.19	0.0035	0.38	0.009	0.97	0.021	0.23	0.0059	0.45	0.022

It is desirable for reliable estimators to have as small values in Table 1 as possible; the smallest values in each column are designated by boldface. Under no contamination, the least squares and MM-estimator turn out to be the

Table 2 Estimated variances of b_0 and b_1 obtained by the nonparametric bootstrap study for $n = 60$. Here, $b = (b_0, b_1)^T$ are estimates of $\beta = (\beta_0, \beta_1)^T$ by means of 9 different estimators of Section 2.2.

		No contam.		Contam. I 10 % of outliers		Contam. I 30 % of outliers		Contam. II 10 % of outliers		Contam. II 30 % of outliers	
		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	LS	**0.051**	**0.0022**	0.254	0.0044	0.279	0.0044	0.167	0.0097	0.951	0.0263
2	S-est.	0.119	0.0054	0.168	0.0034	0.233	0.0030	0.111	0.0059	0.197	0.0252
3	MM-est.	0.070	0.0037	0.114	0.0019	**0.207**	**0.0019**	0.079	0.0036	**0.195**	**0.0050**
4	LTS-0.55	0.055	0.0029	**0.091**	**0.0016**	0.271	0.0026	0.083	**0.0032**	0.280	0.0074
5	LTS-0.80	0.055	0.0024	0.111	0.0017	0.438	0.0049	**0.071**	**0.0032**	0.380	0.0091
6	LWS-A	0.071	0.0037	0.139	0.0025	0.220	0.0025	0.074	0.0047	0.305	0.0088
7	LWS-B	0.115	0.0057	0.164	0.0028	0.269	0.0029	0.114	0.0059	0.215	**0.0050**
8	LWS-C	0.138	0.0061	0.209	0.0037	0.229	0.0028	0.143	0.0060	0.230	0.0048
9	LWS-D	0.054	0.0027	0.170	0.0036	0.279	0.0030	0.098	0.0063	0.600	0.0162

best among all estimators (in terms of the smallest variances). For contaminated data, the following observation is true. If an estimator is particularly good in estimating the slope, then it does not necessarily hold that it is very suitable also for the intercept, and vice versa. We cannot say that any of the four weighting choices is always the best for the LWS. Although it is rather difficult to extract some general conclusions from the results, we can see that the LTS is recommendable only for a smaller contamination. MM-estimators perform well overall, while the LWS estimators (with suitable weights) may be recommendable in some situations with the ability to outperform MM-estimators under a larger contamination.

Results for $n = 60$ are more favorable for MM-estimators, while the LWS performs relatively weaker compared to the case with $n = 30$. To the best of our knowledge, this is a first result of such kind suggesting that the LWS estimator is recommendable especially for smaller sample sizes.

Algorithm 2 Nonparametric bootstrap for the LWS in linear regression.

Input: Data rows $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$
Input: $S > 0$
Output: Empirical covariance matrix computed from individual estimates of $\hat{\gamma}_{LWS}$

1: Compute the least weighted squares estimator $\hat{\beta}_{LWS}$ of β in model (1)
2: **for** $s = 1$ to S **do** // repeat in order to obtain the empirical distribution
3: Generate n new bootstrap data rows

$$({}_{(s)}X^*_{j1}, \ldots, {}_{(s)}X^*_{jp}, {}_{(s)}Y^*_j), \quad j = 1, \ldots, n, \tag{11}$$

by sampling with replacement from data rows $(X_{i1}, \ldots, X_{ip}, Y_i)$, $i = 1, \ldots, n$
4: Consider a linear regression model in the form

$${}_{(s)}Y^*_j = {}_{(s)}\gamma_0 + {}_{(s)}\gamma_{1(s)}X^*_{j1} + \cdots + {}_{(s)}\gamma_{p(s)}X^*_{jp} + {}_{(s)}v_j \tag{12}$$

with $j = 1, \ldots, n$ and random errors $v_1, \ldots, v_n$
5: Estimate ${}_{(s)}\gamma = ({}_{(s)}\gamma_0, {}_{(s)}\gamma_1, \ldots, {}_{(s)}\gamma_p)^T$ in (12) by the LWS
6: Store the estimate from the previous step as ${}_{(s)}\hat{\gamma}_{LWS}$
7: **end for**
8: Compute the empirical covariance matrix from values ${}_{(s)}\hat{\gamma}_{LWS}$, $r = 1, \ldots, R$

4 Conclusions

While it remains important to investigate the robustness of various statistical or econometric tools against outliers or measurement errors [1], this paper is interested in comparing the variability of several highly robust regression estimators. We use a nonparametric bootstrap approach to compare the variability of S-estimators, MM-estimators, the least trimmed squares, and also the (much less popular) least weighted squares.

Let us discuss the results of the numerical studies performed by means of nonparametric bootstrap. If there is

no contamination of the data, the least squares (without surprise) perform the best, while the MM-estimator yields comparable results. If there is a smaller contamination by outliers, good results are obtained with MM-estimators, LWS with linear weights, or sometimes with a suitable LTS. If there is a more severe contamination by outliers, MM-estimators retain the best performance in some situations, but the LWS with logistic or trimmed linear weights are able to overcome MM-estimators in some situations.

On the whole, although MM-estimators are sophistically defined to combine robustness with efficiency, they are definitely far from being always the best, especially for $n = 30$. The LWS estimator has the potential to stay slightly behind the MM-estimator, or even sometimes to overcome its performance. In addition, the LWS estimator (with suitable weights) seems also superior to the much more popular LTS. It seems an interesting novel result that the LWS is able to outperform MM-estimators for $n = 30$ and a potential argument in favor of the LWS estimator. Indeed, recent trends in robust statistics are focused on such smaller samples sizes; we may refer to a recent theoretical paper [17], where attention is paid to the performance of robust estimators for smaller sample sizes.

As future research, the results deserve to be verified for regression models with a larger p. In addition, we would like to use the results as inspiration for proposing highly robust estimation techniques for nonlinear regression models, especially for artificial neural networks.

Acknowledgements

The work was supported by the Czech Science Foundation projects 19-05704S and 17-01251S.

References

[1] Broniatowski, M., Jurečková, J., & Kalina, J. (2018). Likelihood ratio testing under measurement errors. *Entropy*, 20 (12), Article 966.

[2] Davies, L. (1990). The asymptotics of S-estimators in the linear regression model. *Annals of Statistics*, 18, 1651–1675.

[3] Efron, B. & Tibshirani, R.J. (1994). *An introduction to the bootstrap.* Boca Raton: Chapman & Hall/ CRC.

[4] Gschwandtner, M.& Filzmoser, P. (2012). Computing robust regression estimators: Developments since Dutter (1977). *Austrian Journal of Statistics*, 41, 45–58.

[5] Hahn, J. & Ridder, G. (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica*, 81, 315–340.

[6] Jurečková, J., Picek, J., & Schindler, M. (2019). *Robust statistical methods with R.* Second edn. Boca Raton: CRC Press.

[7] Kalina, J. (2012). Facial symmetry in robust anthropometrics. *Journal of Forensic Sciences*, 57, 691–698.

[8] Kalina, J. (2013). Highly robust methods in data mining. *Serbian Journal of Management*, 8, 9–24.

[9] Kalina, J., Vašaničová, P. & Litavcová, E. (2019). Regression quantiles under heteroscedasticity and multicollinearity: Analysis of travel and tourism competitiveness. *Ekonomický časopis/Journal of Economics*, 67, 69–85.

[10] MacKinnon, J. (2002). Bootstrap inference in econometrics. *Canadian Journal of Economics*, 35, 615–645.

[11] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org/.

[12] Rousseeuw, P.J. & Van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery*, 12, 29–45.

[13] Todorov, V. & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32, 1–47.

[14] Víšek, J.Á. (2006). The least weighted squares II. Consistency and asymptotic normality. *Bulletin of the Czech Econometric Society*, 16, 1–28.

[15] Víšek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47, 179–206.

[16] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, 15, 642–656.

[17] Zhou, W.X., Bose K., Fan, J., & Liu, H. (2018). A new perspective on robust M-estimation: Finite sample theory and applications to dependence-adjusted multiple testing. *Annals of Statistics*, 46, 1904–1931.

Forecasting Inflation in Small Open Economies

Jiří Georgiev[1]

Abstract. Inflation forecasts are a crucial part of economic decision-making. The predictive capabilities of multivariate autoregressive models and their non-linear extensions for inflation forecasting on data of selected small open economies within the European Union are examined in this paper. With the introduction of non-linearity and with the growth of dimensions, the number of estimated coefficients is growing rapidly. Therefore, the models are more flexible and have the potential to capture even more complex relationships, but the risk of overfitting is increasing. The predictive performance of the vector autoregressive model (VAR), Factor-Augmented VAR (FAVAR) model and feedforward neural network estimated with various settings and different regularization approaches are compared. Simple autoregressive model is used as a benchmark and errors are calculated by rolling windows with short forecasting horizon.

Keywords: VAR, FAVAR, feedforward neural network, inflation

1 Introduction

Inflation is a key economic statistic for monetary authorities and other economic subjects. Inflation forecast aid in decision-making and policy forming. During the last decade, world economies underwent significant changes after the world economic crisis. Thus, previously successful forecasting approaches may need re-evaluation. Forecasting capabilities can be improved by expanding the classical linear univariate model to multivariate and allowing non-linear relationships. In the case of inflation, univariate models cannot capture dynamic interactions between different macroeconomic variables. This causes the misspecification of the model and creates one of the possible causes of forecasting errors [1]. The main drawback of multivariate models is quickly increasing number of parameters to estimate due to increasing number of variables. Various studies, for instance Marcellino [2], Nakamura [3], Aiken [4], Choudhary [5], have considered the usage of neural networks for inflation forecasting. Most of these studies focused only on univariate models and their results suggest that neural network can perform the same or better than their linear alternatives, ARIMA models. Several studies, e.g. Bernanke [6], Stock and Watson [7], Artis [8], have shown that factor-augmented vector autoregressive model (FAVAR) outperforms classical vector autoregressive model (VAR) in settings where a large number of variables is available.

This paper aims to carry out a comparison of multivariate time series models for inflation forecasting across selected small open economies. Considered multivariate time series models are vector autoregressive model, factor-augmented vector autoregressive and feedforward neural network. The main goal is to evaluate whether an introduction of non-linearity will be beneficial for forecasting and assess the importance of regularisation as a measurement to prevent overfitting. Another goal is to determine whether a reduction of dimensionality with factor-model approach can aid forecasting in terms of preventing overfitting even in case of models with a low number of variables.

The structure of this paper is as follows. Used models, datasets and methodology for evaluation of forecasting capabilities are described in section 2. Results are presented in Section 3, and the paper is concluded in Section 4.

2 Models, Dataset and Methodology

2.1 Models

Vector autoregressive model

VAR model is a popular and flexible model, that is well suited to capture the dynamic behaviour of multivariate time series, but often needs to be expanded by exogenous variables. The exogenous variable can be a deterministic trend, seasonal dummy variables or any other variables, that can be useful for forecasting. This extended p-lag VAR(p) model can be defined as follows

$$\boldsymbol{Y}_t = \boldsymbol{c} + \boldsymbol{\Pi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Pi}_2 \boldsymbol{Y}_{t-2} + \cdots + \boldsymbol{\Pi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{G} \boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

[1] University of Economics, Prague, Department of Econometrics, jiri.georgiev@vse.cz

where $\boldsymbol{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is $(n \times 1)$ vector of time series variables, $\boldsymbol{D}_t$ is $(l \times 1)$ matrix of deterministic variables, $\boldsymbol{X}_t$ is $(m \times 1)$ matrix of stationary exogenous variables and $\boldsymbol{\Phi}$, $\boldsymbol{G}$, $\boldsymbol{\Pi}$ are parameter matrices [9]. To select appropriate lag length information criterion can be used. Most commonly used criterion are Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ). AIC tends to overestimate and BIC tends to underestimate the number of lags needed [10]. Therefore, BIC might be more appropriate for forecasting applications as underestimation prevents overfitting. Lütkepohl [11] provides a detailed description of information criterion and their usage.

Factor-augmented vector autoregressive model

FAVAR model includes low-dimensional factors in order to capture complex fluctuations or theoretical concepts, that cannot be easily represented by one or two time series, but rather by a wide range of variables. The factor-model approach reduces dimensionality and therefore allows higher dimensional input data. Basic FAVAR model can be defined as

$$\begin{bmatrix} \boldsymbol{F}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \Phi(L) \begin{bmatrix} \boldsymbol{F}_{t-1} \\ \boldsymbol{Y}_{t-1} \end{bmatrix} + v_t, \tag{2}$$

where $\boldsymbol{Y}_t$ is a vector of observable variables and $\boldsymbol{F}_t$ is $K \times 1$ vector of unobserved factors, where K is small and $\Phi(L)$ is a conformable lag polynomial of finite order q. Equation (2) cannot be estimated directly because of unobservable factors $\boldsymbol{F}_t$. It is assumed that observed time series $\boldsymbol{X}_t$ are related to unobserved $\boldsymbol{F}_t$ and to observed $\boldsymbol{Y}_t$ in the following form

$$\boldsymbol{X}_t = \boldsymbol{\beta}^f \boldsymbol{F}_t + \boldsymbol{\beta}^y \boldsymbol{Y}_t + \boldsymbol{e}_t, \tag{3}$$

where $\boldsymbol{e}_t$ is a vector of idiosyncratic error term with zero mean. There are two main approaches to estimate $\boldsymbol{F}_t$, two-step principal components approach [12] and single-step Bayesian likelihood approach. Bernanke [13] shows both methods yield similar results. This contribution considers only computationally less expensive two-step approach. Firstly, observed $\boldsymbol{X}_t$ are used to estimate series of factors $\boldsymbol{F}_t$ with principal component method [7] and then model (2) can be estimated with $\hat{\boldsymbol{F}}_t$. The appropriate number of factors can be selected by scree plot or by evaluation of percentages of cumulative variance. Other approaches based on cross-validation may be considered [14] for applications with longer time series.

Feedforward neural network

A feedforward neural network is a type of deep learning model. This type of neural network consists of layers of interconnected nodes, specifically one input layer, one output layer, and several hidden layers. As the name suggests, information flow only in one direction from the input layer to the output layer and there are no connections in which output is fed back into the model. Output of neuron in k-th hidden layer is given by activation function F,

$$\boldsymbol{h}^{(k)} = F(\boldsymbol{a}^{(k)}), \tag{4}$$

in which $\boldsymbol{a}^{(k)}$ is a weighted sum of

$$\boldsymbol{a}^{(k)} = \boldsymbol{b}^{(k)} + \boldsymbol{W}^{(k)} \boldsymbol{h}^{(k-1)}, \tag{5}$$

where $\boldsymbol{W}^{(k)}$ is a matrix of weights and $\boldsymbol{b}^{(k)}$ is a vector of biases. Output for the network with depth l is $\hat{\boldsymbol{y}} = \boldsymbol{h}^{(l)}$. Mean square error can be used as loss function $L(\hat{\boldsymbol{y}}, \boldsymbol{y})$ and regularisation term may be added. For more detailed description of the structure of the feedforward neural network, see Goodfellow [15]. Most common activation functions are softmax, relu, sigmoid and linear. Neural networks in this contribution are built with `Keras`[1] and `TensorFlow`[2] back-end and gradient-based algorithm ADAM[16] is used for optimisation.

Autoregressive model

Simple autoregressive model, which is used as a benchmark model, can be described as

$$y_t = \beta_0 + \sum_{l=1}^{p} \beta_l y_{t-l} + \gamma_1 x_t + u_t \tag{6}$$

where x_t is exogenous variable, p is a number of lags selected by Schwarz-Bayesian information criterion.

[1] High-level neural network API for R `https://keras.rstudio.com/`

[2] Open source machine learning framework `https://www.tensorflow.org/`

2.2 Dataset

The dataset used in this contribution has been obtained from OECD database[3] with the aid of R package OECD[4]. The analysis focuses on seven small open economies: Austria (AUT), Czech Republic (CZE), Hungary (HUN), Poland (POL), Slovakia (SVK), Sweden (SWE) and Netherlands (NLD). Predictions are made for Customer Price Index *cpi_q* measured as Harmonised Index of Consumer Price (HICP). In multivariate models Gross Domestic Product *gdp_q*, 3 month interest rate *int*3_*q*, Net trade *net_q*, Unit labour costs *ulc_q*, Unemployment rate *unemp_q*, Employment rate *emp_q*, M1 monetary aggregates *m*1_*q*, M3 monetary aggregates *m*3_*q*, Active population *act_q*, Export *exp_q* and Import *imp_q* are used to improve forecasting power. HICP aggregated for 28 states of the European Union is used as an exogenous variable.

To assess whether the series is stationary or nonstationary with unit-root Augmented Dicker Fuller test has been used. For stacionarization first differences have been used. All obtained time series are quarterly, spanning over the period from 1996Q2 to 2018Q1. Due to missing values, not all series are usable for all countries and starting points among countries differ. Overview of available series for each country and their starting point is in table 1.

Country	Start	cpi_q	act_q	emp_q	exp_q	gdp_q	imp_q	int3_q	net_q	ulc_q	unemp_q	m1_q	m3_q
AUT	1996Q2	1	1	1	1	1	1	1	1	1	1	x	x
CZE	1998Q1	1	1	1	1	1	1	1	1	1	1	1	1
HUN	1999Q2	1	1	1	1	1	1	x	1	1	1	1	1
NLD	2000Q2	1	1	1	1	1	1	1	1	1	1	x	x
POL	2002Q1	1	1	1	1	1	1	1	1	1	1	1	1
SVK	1996Q2	1	1	1	1	1	1	x	1	1	1	x	x
SWE	2001Q2	1	1	1	1	1	1	1	1	1	1	1	1

Table 1 Available macroeconomic time series of selected countries

2.3 Methodology

Forecast evaluation is carried out using Mean Square Error (MSE) on "rolling forecast origin" with short forecasting horizon. Rolling forecast origin is a type of cross-validation technique for time series and is implemented as described by Hyndman [17]. Therefore, each model is estimated 10 times, the origin is moved each time by one and MSE is calculated for forecasting horizon with a length of 3. Then final error (MSE_R) is calculated as the average of all MSE errors, where MSE is defined as

$$MSE_h = \frac{1}{h}\sum_{i=1}^{h}(e_i)^2, \tag{7}$$

where h is the length of the horizon and e_i is prediction error for i-th observation. Because MSE_R is aggregated from multiple forecasting errors, it should be a robust measurement of forecasting ability of assessed models and therefore suitable for their comparison. Although for longer time series, it would be possible to further split data into test set used for selecting meta parameters and validation set used for error evaluation.

For each, country all possible combinations of variables, consisting of up to four time series, have been estimated firstly as VAR models. Appropriate lags length has been selected by AIC, BIC, HQ and for each of selected lags, a model has been estimated. Contemporaneous values of European Union inflation rates have been used as an exogenous variable in each estimated model. Models have been estimated for each combination of deterministic variables such as trend, seasonal dummy and dummy for the economic crisis in 2008 as well. Similar approach has been used for FAVAR model. Additionally, the number of components for FAVAR has been selected by a threshold of 90%, 80% 70% of cumulative variance. Due to computationally demanding nature of neural networks, only selection of variable combinations has been estimated. Only combinations that proved effective in VAR and FAVAR models have been used. Different sets of meta parameters, such as number of hidden layers (ranging from none to 4), number of hidden neurons, activation functions (linear, softmax, relu), regularization term with λ, early stop method [18] were used for all estimated combinations. Sets of used meta parameters vary among countries, but a minimum of 2300 models have been estimated for each country. All networks are estimated with a maximum of 4000 epochs and linear output layer activation function. In total 11616 VAR models, 40846 FAVAR models and 19594 neural nets have been estimated.

[3] OECD database -https://data.oecd.org/

[4] https://cran.r-project.org/web/packages/OECD/OECD.pdf

3 Results

This section summarizes the results of the forecasting exercise. Table 2 reports the lowest forecasting errors for each country and each model. Bold entries represent the models with the lowest MSE_R for a given country. Results for best performing neural networks, VAR models and FAVAR models are presented in tables 3, 4 and 5 respectively. All models, with the exception of neural networks for the Netherlands and some neural networks

Country	Neural Network (without regularization)	Neural Network (with regularization)	VAR	FAVAR	Benchmark model
AUT	0.0544	**0.0411**	0.0606	0.0684	0.0688
CZE	0.1078	0.0858	**0.0576**	0.0662	0.0931
HUN	0.1175	0.1172	0.0935	**0.0934**	0.2637
NLD	0.1108	0.0566	**0.0236**	0.0351	0.0472
POL	0.1190	0.1241	**0.0890**	0.1142	0.1539
SVK	0.0888	0.0686	**0.0450**	0.0522	0.1391
SWE	0.1477	**0.0312**	0.0507	0.0525	0.2118

Table 2 MSE computed on rolling origin forecast

without regularisation, significantly improved forecast in comparisons to the benchmark model. The results suggest that regularisation was a key feature for the successful performance of a neural network. Only for Poland, neural network without regularisation performed better than the neural network with regularisation, but early stop method was used in this network. Interestingly, best performing neural network for Sweden was a network without any hidden layer and therefore similar in structure to VAR model but estimated with regularisation. Furthermore, the network for Austria contains only *cpi_q* and exogenous variables. This network shows that some countries might utilise the introduction of nonlinearity more than additional endogenous variables.

Results in table 5 show that FAVAR reduced dimensionality because the number of used factors K is in all cases lower than the number of original variables at least by one. Although this reduction improved predictions only for one country, FAVAR results are similar to VAR results.

Country	Variables	Activation function	Regularization	λ	Early Stop	Hidden units	Hidden layers	Number of lags	MSE_R
AUT	cpi_q	softmax	True	0.0100	False	50	3	4	0.0411
CZE	cpi_q,imp_q,unemp_q	softmax	True	0.0005	False	20	4	4	0.0858
HUN	cpi_q,act_q,unemp_q	linear	True	0.0010	False	55	1	1	0.1172
NLD	cpi_q	linear	True	0.0001	False	15	3	5	0.0566
POL	cpi_q,gdp_q,ulc_q	linear	False	x	True	30	2	6	0.1190
SVK	cpi_q,ulc_q	linear	True	0.0010	False	45	3	5	0.0686
SWE	cpi_q,act_q,gdp_q,imp_q	linear	True	0.1000	True	x	0	3	0.0312

Table 3 Results for best-performing feedforward neural networks

Country	Variables	Lag	MSE_R
AUT	cpi_q,act_q,imp_q	1	0.0606
CZE	cpi_q,gdp_q,imp_q	4	0.0576
HUN	cpi_q,act_q,unemp_q	1	0.0935
NLD	cpi_q,gdp_q,int3_q	4	0.0236
POL	cpi_q,gdp_q,ulc_q	5	0.0890
SVK	cpi_q,act_q,ulc_q	4	0.0450
SWE	cpi_q,gdp_q,int3_q,ulc_q	4	0.0507

Table 4 Results for best performing VAR

Country	Variables	K	Lag	MSE_R
AUT	cpi_q,act_q,imp_q,int3_q,unemp_q	2	1	0.0684
CZE	cpi_q,gdp_q,m1_q,unemp_q	1	4	0.0662
HUN	cpi_q,act_q,imp_q,unemp_q	2	1	0.0934
NLD	cpi_q,gdp_q,imp_q,int3_q,unemp_q	1	5	0.0351
POL	cpi_q,gdp_q,imp_q,m1_q,unemp_q	1	5	0.1142
SVK	cpi_q,act_q,imp_q,unemp_q	1	6	0.0522
SWE	cpi_q,ulc_q,unemp_q	1	5	0.0525

Table 5 Results for best performing FAVAR

4 Conclusion

The goal of this contribution was to evaluate the predictive performance of VAR model, FAVAR model estimated by the two-step approach and feedforward neural network for inflation in small open economies. To do so, models with various settings and variables were estimated on data for seven small open economies and rolling origin forecast error was computed. With some exceptions, selected models performed better than the benchmark model. Among these models, VAR model performed the best, followed by FAVAR model. An interesting finding is that feedforward neural networks which are able to capture more complex and nonlinear relations performed relatively poorly. Because searching space of possible settings for neural networks is computationally expensive, a possibility that there are better settings cannot be ruled out. FAVAR managed to reduce the dimensionality of original data but failed to do so in a way that would be significantly beneficial to the performance of inflation forecasting.

Acknowledgements

Supported by the grant No. IGA F4/53/2019, Faculty of Informatics and Statistics, University of Economics, Prague

References

[1] D. F. Hendry and M. P. Clements, "Economic forecasting: Some lessons from recent research", *Economic Modelling*, vol. 20, no. 2, pp. 301–329, Mar. 2003, ISSN: 02649993.

[2] M. Marcellino, "Instability and non-linearity in the EMU", *IEP-Università Bocconi and CEPR*, 2002.

[3] E. Nakamura, "Inflation forecasting using a neural network", *Economics letters*, 2004.

[4] M. Aiken, "Using a neural network to forecast inflation", *Industrial Managment & Data Systems*, 1999.

[5] M. A. Choudhary and A. Haider, "Neural network models for inflation forecasting: An appraisal", *Applied Economics*, vol. 44, no. 20, pp. 2631–2635, Jul. 2012.

[6] B. S. Bernanke and J. Boivin, "Monetary policy in a data-rich environment", *Journal of Monetary Economics*, Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information, vol. 50, no. 3, pp. 525–546, Apr. 1, 2003, ISSN: 0304-3932.

[7] J. H. Stock and M. W. Watson, "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business & Economic Statistics*, vol. 20, no. 2, pp. 147–162, Apr. 2002, ISSN: 0735-0015, 1537-2707.

[8] M. J. Artis, A. Banerjee, and M. Marcellino, "Factor forecasts for the UK", *Journal of Forecasting*, vol. 24, no. 4, pp. 279–298, 2005, ISSN: 1099-131X.

[9] E. Zivot and J. Wang, *Modeling Financial Time Series with S-PLUS®*. Springer New York, Oct. 10, 2007.

[10] M. Gredenhoff and S. Karlssony, "Lag-length Selection in VAR-models Using Equal and Unequal Lag-Length Procedures", *Working Paper Series in Economics and Finance*, 1997.

[11] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

[12] J. H. Stock and M. W. Watson, "Forecasting Using Principal Components From a Large Number of Predictors", *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1167–1179, Dec. 2002, ISSN: 0162-1459, 1537-274X.

[13] B. S. Bernanke, J. Boivin, and P. Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", *Quarterly Journal of Economics*, p. 36,

[14] J. Josse and F. Husson, "Selecting the number of components in principal component analysis using cross-validation approximations", *Computational Statistics & Data Analysis*, vol. 56, no. 6, pp. 1869–1879, Jun. 1, 2012, ISSN: 0167-9473.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, Jan. 3, 2017, 800 pp., ISBN: 0-262-03561-8.

[16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", Dec. 22, 2014. arXiv: 1412.6980v9.

[17] R. J. Hyndman, "Measuring forecast accuracy", 2014.

[18] L. Prechelt, *Neural Networks: Tricks of the Trade (Lecture Notes in Computer Science)*. Springer, 2012, ISBN: 978-3-642-35289-8.

European insurance market analysis via functional data clustering techniques

Stavros Athanasiadis[1], Tomáš Mrkvička[2]

Abstract. Insurance penetration as a high-level indicator of an insurance market's development exhibits significant variations over time across countries. Cross-country comparisons on the link between insurance and economic growth are better off when insurance development homogeneity is present. Motivated by this evidence, this study is aimed to provide a data-driven and meaningful clustering of European countries in terms of their insurance penetration rates that are considered as functions (curves). The ultimate goal is the extraction and visualization of the representative curves that characterize the homogeneous clusters of European insurance market. To this end, we apply functional data clustering methods that fall into three major categories: distance-based methods, filtering methods, and adaptive methods. The data consist of insurance penetration rates sampled from 34 European countries and observed between 2004 and 2016; that is before, during and post-financial and sovereign debt crises. Our results - the clusters- are analyzed from a qualitative point of view, detecting visually whether they are distinguishable and preserve the magnitude and shape of the cluster member curves.

Keywords: Insurance, Insurance penetration, Time series clustering, Functional data analysis, Functional data clustering

JEL Classification: C22, C38, G22
AMS Classification: 91B30, 91C20, 62H30, 62P05

1 Introduction

The recent financial crisis illustrated the need to recognize the systemic risk in insurance and to realize the extent of its effects on the real economy by investigating the link between the development of insurance markets and economic growth. Insurance development level is usually measured by insurance penetration (IP) rate i.e. the ratio of insurance premiums to the gross domestic product (GDP) [21]. We observe European insurance markets outlook on an annual basis for collecting IP rates across different countries over yearly periods (2004 to 2016).

We propose representing the time evolution of IP rates data from each country first by time series data and then by functional data. The conversion of time series data into functional data brings to light the IP curves that carry the historical information about the behavior of insurance development in certain countries. Often, cross-country comparisons on the causal effects of economic growth on insurance [8] or on the causal effects of insurance on economic growth [1] face the consequences of the insurance development heterogeneity across countries [18]. Consequently, it is necessary to find homogeneous groups of countries based on their IP curves to improve the quality of such causal effect analysis.

The aim of this study, therefore, is to identify meaningful and homogeneous clusters of European countries according to two main properties of their IP curves, namely, the magnitude (size) and the shape. In other words, we want to cluster IP curves by their shape, giving also attention to fluctuations or to the total magnitude of the curves. In [2], authors' attempt showed that multivariate clustering methods on IP rates were unable to extract meaningful clusters of countries, in the sense that the cluster representative IP pattern cannot reveal appropriately both the magnitude and the shape of the cluster member IP profiles at the same time. The results came as no surprise, since in multivariate situation clustering methods do not take into account for the high dimensionality, the time shifting, the non-linear high correlation and the continuous change behavior of the IP rates over time.

In light of the difficulties encountered by the multivariate clustering methods, we propose to cluster the IP rates by means of either time series data clustering methods (based on dissimilarity measures) or functional data

[1] University of South Bohemia, Department of Applied Mathematics and Informatics, Studentská 13, 37005 České Budějovice, Czech Republic, e-mail: athans00@ef.jcu.cz.
[2] University of South Bohemia, Department of Applied Mathematics and Informatics, Studentská 13, 37005 České Budějovice, Czech Republic, e-mail: mrkvicka.toma@gmail.com.

clustering methods. With respect to time series clustering methods, only the K-medoids with an autocorrelation-based distance between time series has been applied. Whereas, with respect to functional data clustering methods, the K-means clustering on functional principal components (FPC) scores, the discriminative functional mixture model, and the K-medoids with semi-metric distance between IP curves have been applied. In our effort to do clustering based on both the magnitude and the shape of the IP curves, we choose to forward first the raw IP curves and then a rescaled version of these curves - the centralized IP curves - to functional clustering methods. Further, we assess the performance of all clustering methods by visually detecting their generated clustering solutions. Particularly, we detect whether no overlapping clusters are present and whether clusters recover the shape information of cluster member IP curves.

2 Methods

We used non-missing time series data at country level taken at fixed year-end points from Swiss Re Sigma database (2016) [24]. This cluster analysis concentrated on total (life and non-life insurance) IP rates of European countries during the period between 2004 and 2016. The data available consisted of 13 time series percentage-valued samples of 34 European countries. We followed two clustering procedures for European insurance market segmentation. One conducted clustering methods on time series data (IP time series) and the other conducted clustering methods after converting the time series data into functional data (IP curves).

The functional clustering methods are based on both the magnitude and shape of the IP curves. Particularly, the shape information within data is uncovered by initial data transformation. In line with [6], we rescaled each raw IP curve by subtracting its mean value from the curve itself, resulting in the centralized IP curves. This operation was intended to mitigate the widely different magnitudes of the IP curves. By shifting each curve towards its center, the overall shape of each IP curve could be better identified without the complication of comparing IP curves with different magnitudes. After this, clustering was performed first in terms of magnitude, on raw IP curves, and then it was performed in terms of shape, on centralized IP curves. In the latter case, the results were converted back to the raw IP curves, for their graphical presentation and analytical interpretation.

The first clustering procedure consisted of utilizing distance-based time series clustering methods. In time series data, each observation in the set $\{Y_{1t}, \dots, Y_{nt}\}$ represents a time series i.e. a sequence of realizations of a real-valued random variable $Y = \{Y_t, t \in \mathbb{N}\}$ over discrete time $t \in \{1, \dots, T\}$, where n is the number of time series under consideration. This study is concerned with the so-called model-free dissimilarity measures [19]. More precisely, we turned to the calculation of every pairwise autocorrelation-based dissimilarity d_{ACFG}, [10] between two IP series with a maximum delay of $L = 12$ and geometrically decaying weights with $p = 0.5$. A pairwise dissimilarity matrix D is first created out of these dissimilarities and then forwarded to the K-medoids method (partitioning around medoids (PAM) [15]). We should note that before calculating d_{ACFG}, an autocorrelation representation of time series has been considered to remove time shifts and highlight the shape information within data.

The second clustering procedure consisted of transforming time series data to functional data during clustering. In functional data, each observation in the set $\{x_1(t), \dots, x_n(t)\}$ represents an independent sample path of a stochastic process $X = \{X(t), t \in [0, T]\}$ over continuous time $t \in [0, T]$, where n is the number of sample paths under consideration. In practice, sample paths or curves $x_i(t), i = 1, \dots, n$ are observed in a finite set of time points $\{t_{i0}, \dots, t_{i,T}\}$ so that we are able to observe only the discrete data and not the continuous functions. A classical approach to recover the functional nature of the data is through a finite basis expansion. In this study, to estimate this expansion, we used a smoothing procedure based on cubic B-spline functions [22] of order 4.

In what follows, we review briefly the functional data clustering methods that are summarized in [14] and divided into three major categories: distance-based methods, filtering methods, and adaptive methods. We start with the distance-based functional clustering, which resembles the distance-based time series clustering. However, they exploit distance measures specifically developed for curves and defined in terms of integral of either raw curves or first derivative of raw curves [13]. In this study, we focused on the integrals of raw and centralized IP curves (rather than on the integral of first derivative of raw IP curves) that were approximated by the Trapezoid rule [4]. We were able then to measure the distance $d0$ between pairs of raw or centralized IP curves. For clustering analysis purposes, we used again the K-medoids (PAM) method with the newly defined distance metric $d0$.

Proceeding with the filtering methods, we present functional principal component analysis (FPCA) [20], which is a generalization of ordinary (multivariate) principal component analysis (PCA) [12]. Similar to PCA, FPCA provides common principal components (PCs) $\xi_i(t), i = 1, 2, \dots$ that give the directions of sample curves, while the random scores z_i, represent the coordinates of the projected sample onto the i^{th} direction. Since the number of directions that explain the most variation of the data is usually much lower than the number of sample curves,

FPCA is usually applied for dimensionality reduction. The aim of FPCA is, therefore, to find the dominant modes of a set of curves or equivalently the eigenfunctions $\xi_i(t)$ that maximize the variance function [11] of the PC scores z_i, subject to the constraints that the eigenfunctions need to form an orthonormal system. In this study, after having conducted FPCA on the IP curves, we applied K-means [16] on the PC scores of these curves from the first few PCs that explained a vast percentage (over 90%) of the variation of the curves.

Finally, we close up this review with the description of one of the most recent adaptive clustering methods, which utilizes a model-based clustering method proposed by [3] and is called funFEM method. In brief, the coefficients ω_i of the B-spline decomposition of each curve $x_i(t)$, where $i = 1, \dots, n$, are assumed to follow a K component Gaussian mixture distribution as defined in (1). The expectation maximization (EM) algorithm [17] is enhanced with an additional step in order to estimate the parameters of Gaussian components (clusters) in the mixture. The EM estimates these parameters in maximum likelihood sense and in a discriminative functional subspace of lower fixed dimension $d < K$ (common over all K) than the original one p.

$$P(\omega) = \sum_{\kappa=1}^{K} \pi_\kappa \phi(\omega; \eta\mu_k, \eta^\tau \Sigma_\kappa \eta + \Lambda) \quad (1)$$

P denotes the marginal distribution, π_κ is the mixing probability of the k^{th} cluster, ϕ is the standard Gaussian density function, μ_k and Σ_κ are the mean and covariance matrix of the k^{th} cluster for the mapping of coefficient ω into the discriminative subspace, where η is a matrix representing the mapping to the discriminative subspace. Additionally, Λ denotes the covariance matrix related to measurement noise, which is modelled through a family of 12 different discriminative functional models (DFM) summarized also in [3]. In this study, we performed funFEM method with K-means initialization method.

For heuristic clustering methods (K-means and PAM), we estimated the optimal number of clusters using the results of NbClust package [5]. This R package provides the results of 30 validity indices that estimate the number of clusters in a dataset. It provides also the majority vote of indices in terms of the best number of clusters. The final choice of the optimal number of clusters was also subject to our expert judgement. As for funFEM method, we determined which DFM model to use and the optimal number of components (clusters) K, by using the Bayesian information criterion (BIC) [23]. The number of clusters that stabilizes BIC was chosen to be the optimal one.

3 Research results

Table 1 shows the selected cluster number K for each proposed clustering method along with the criteria consulted before forming our selection. Given the cluster number, K, we conducted four clustering methods: PAM clustering using the d_{ACFG} dissimilarity measure (PAM-d_{ACFG}), K-means clustering on functional PC scores (Kmeans-FPCscores), Gaussian mixture model-based functional clustering (funFEM) and PAM method with semi-metric distance $d0$ (PAM-$d0$). For each clustering method, we used the available codes implemented in libraries of the R (version 3.5.3) programming language such as TSclust, fda.usc, funFEM and stats among others.

Clustering Method	Criterion	Selected Cluster Number
PAM-d_{ACFG}	30 Indices in NbClust	4
Kmeans-FPCscores	30 Indices in NbClust	3
funFEM	BIC	3
PAM-$d0$	30 Indices in NbClust	3

Table 1 Cluster Number Selection

We conducted two clustering tests in order to test the ability of the proposed functional data clustering methods to cluster properly IP curves of similar magnitude and shape. First, we did functional data clustering on raw IP curves (magnitude-based clustering test) and then on centralized IP curves (shape-based clustering test). We should note that the raw IP curves were transformed to centralized IP curves (used in all shape-based clustering tests) and then back-transformed to the raw IP curves before presenting the results. The proposed time series clustering method is a shape-based clustering, since the shape heterogeneity of IP time series became more transparent after having used an autocorrelation representation of IP time series in the initial step of this clustering method.

Then, we drew cluster representative IP curves (rep. IP curves) and cluster curve plots for each clustering test. These plots enabled us to visually analyze the performance of the clustering methods in terms of precision in allocating IP curves with both similar magnitude and shape into the same cluster. Since the financial and sovereign debt crises in Europe began in the middle of 2007 and ended at the end of 2011, we depicted in our figures this

period by vertical dotted lines that intersect the axis of years to 2007 and 2011. In this way, we implicitly constructed three periods of different economic conditions within the period 2004-2016. These periods are called phase-period A, B and C and correspond to the period before, during and post-financial and sovereign debt crises. At the end, to characterize the clusters of the IP curves, we chose the clustering method that generated the best clustering solution. The results are summarized below.

Figure 1 presents the rep. IP curves generated by PAM-d_{ACFG} method. We observe that clusters did not recover the magnitude of their cluster member IP curves, confirming that only shape-based clusters are generated by this method. Figures 2, 3, 4 and 5 show that the clusters did not represent completely the shape variations of the cluster member IP curves. The bottom line is that the shape of IP profiles was not perfectly captured by the autocorrelation representation of IP time series (time shift removal) that was employed before performing PAM-d_{ACFG} clustering.

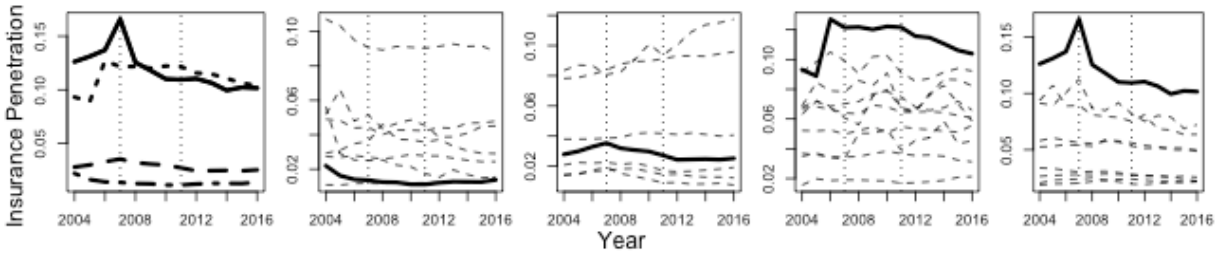


Figure 1 rep. IP curves **Figure 2** Cluster 1 **Figure 3** Cluster 2 **Figure 4** Cluster 3 **Figure 5** Cluster 4

Figures 7, 8 and 9 plot the three obtained clusters from the funFEM method on raw IP curves. In Figure 6, the cluster rep. IP curves are shown to have random fluctuations spanned over not overlapping IP levels. As it turns out, clusters have recovered the magnitude of the cluster member IP curves, but not their shape.

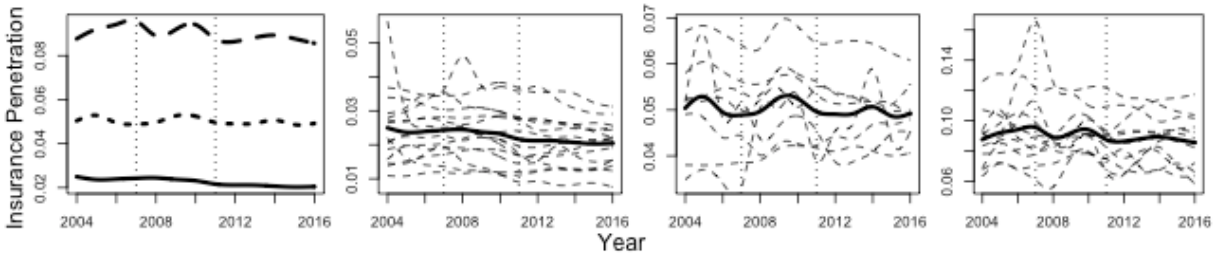


Figure 6 rep. IP curves **Figure 7** Cluster 1 **Figure 8** Cluster 2 **Figure 9** Cluster 3

Figures 11, 12 and 13 point that funFEM method on centralized IP curves identified three shape-based clusters. In Figure 10, Cluster 1 (dotted curve) has a stable shape, Cluster 2 (long dashed curve) has a varying and increasing shape and Cluster 3 (solid curve) has a light varying and decreasing shape (after the peak occurred in year 2007) that eventually becomes flatter towards the right end. We see that many increasing or decreasing shape IP curves of Cluster 2 or 3, respectively, are misallocated to Cluster 1 (clustered as stable shape curves). This misallocation might be resulted from the fact that in a functional data context the curve distribution notion is not well-defined [7].

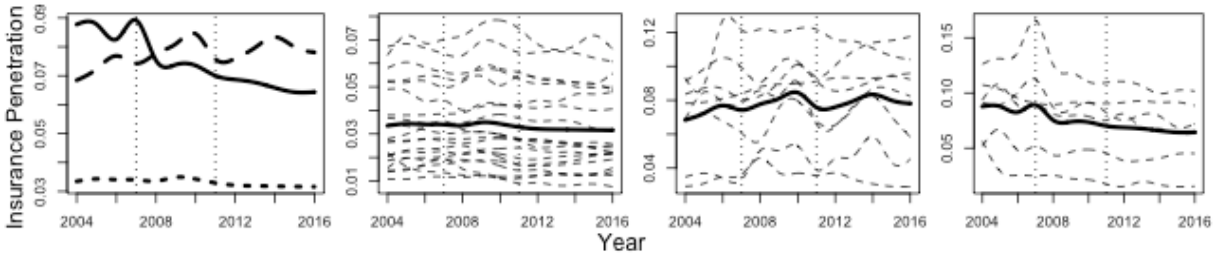


Figure 10 rep. IP curves **Figure 11** Cluster 1 **Figure 12** Cluster 2 **Figure 13** Cluster 3

The performance results of the Kmeans-FPCscores method on raw and centralized IP curves is almost similar to that of funFEM method. However, they exhibited different degrees of misallocation of centralized IP curves to clusters: Kmeans-FPCscores had a higher degree of misallocation compared to funFEM method (confirmed by visual inspection of curve plots). This might be an evidence that not all clusters share the same number of common features (PCs) that extracted from FPCA.

Figures 15, 16 and 17 plot the clustering solution generated by the PAM-$d0$ method on raw IP curves. As we can see, the clusters have adequately summarized the information carried in the cluster member IP curves in terms of their magnitude, but not in terms of their shape. Figure 14 shows the selected cluster rep. IP curves that exhibited few random fluctuations over time, but their IP levels remained distinct without overlapping.

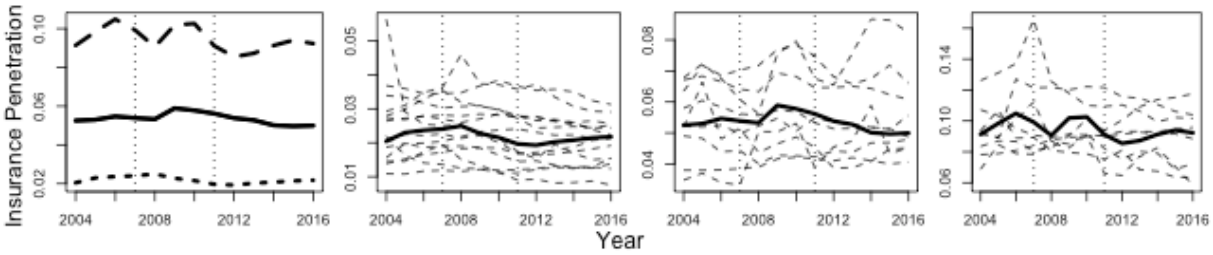


Figure 14 rep. IP curves **Figure 15** Cluster 1 **Figure 16** Cluster 2 **Figure 17** Cluster 3

Figures 18, 19, 20 and 21 visualize the clustering of PAM-$d0$ method on centralized IP curves. Despite the formation of 3 clusters by this method, we faced once again the shape-based misallocation of the IP curves that observed in the results of the other two alternatives of functional data clustering. This might be because the shape information within data was too complex that could not be efficiently modeled by $d0$ distance measure [14].

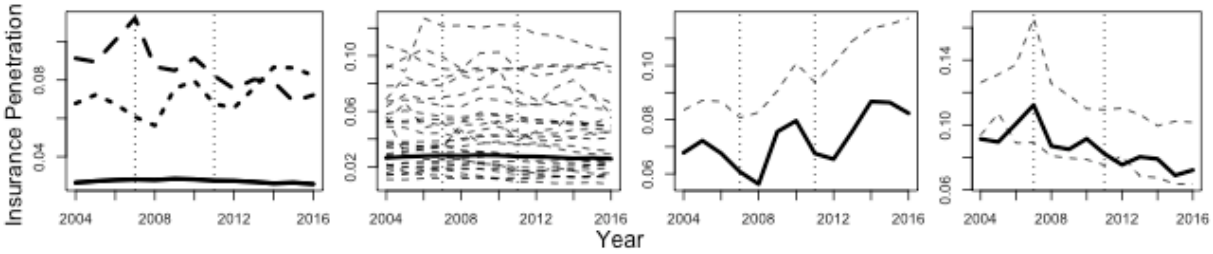


Figure 18 rep. IP curves **Figure 19** Cluster 1 **Figure 20** Cluster 2 **Figure 21** Cluster 3

As performance results conclusion, we can state that both Kmeans-FPCscores and funFEM performed fairly well with the two datasets of IP curves compared to PAM-d_{ACFG} and PAM-$d0$. However, the clustering solution of funFEM method on raw IP curves was our choice for the characterisation of clusters. As seen in Figure 6, Cluster 1 (solid curve) has a downward trend which becomes more evident and rather steep throughout the phase-period B. Both Clusters 2 (dotted curve) and 3 (long dash curve) fluctuate with random cycles of ups and downs during all phase periods. However, Cluster 3 fluctuates more heavily than Cluster 2 over high ranked IP levels. Overall, these figures indicate that Cluster 3 unites the well-developed insurance markets with their unstable behavior, Cluster 2 the medium-developed insurance markets that managed to follow closely medium ranked IP levels, and Cluster 1 the less-developed insurance markets that faced shrink challenges especially during financial and sovereign debt crises period.

4 Conclusions

This study is carried out in two perspectives. The first is to cluster European countries according to their IP rates and provide a cluster characterization (the best possible) that reflects the over-time IP pattern of countries within this cluster. To this end, we applied four clustering methods that treat IP rates data as time series data or as functional data during clustering. These methods are PAM-d_{ACFG}, Kmeans-FPCscores, funFEM and PAM-$d0$. The second is to assess whether these clustering methods are producing cluster rep. IP curves that retain, at the same time, the magnitude and the shape properties of the cluster member IP curves. This assessment is accomplished by applying the clustering methods not only on raw IP curves (on magnitude basis), but also on centralized IP curves (on shape basis).

Under the first perspective, on one side appeared the well-developed insurance markets that faced heavy fluctuations in their IP levels. On the other side appeared the less-developed insurance markets with their even more deteriorated IP levels after the financial and sovereign debt crises. Finally, in the middle appeared the medium-developed insurance markets that retained expectations for stable medium ranked IP levels. Under the second perspective, we succeeded to do clustering that drew quite good distinction between clusters of countries as long as the magnitude of their IP curves were important. Clustering became more difficult when we focused on shape of the IP curves, so that the methods performed not as well as expected to figure out the underlying shapes.

Consequently, we demonstrated that the selected clustering methods were unable to produce meaningful clusters of countries that could reveal, at the same time, both the magnitude and the shape of IP curves. We also found that the proposed clustering methods suffer from certain difficulties. For distance-based time series clustering, the difficulty is caused by the uncertainty of feature-based representations of original data (time series) such as autocorrelations. For distance-based functional clustering, the complex shape mechanisms underlying the data may not be well understood by the developed distance measure. For filtering methods, the FPCA approach may form clusters that do not share the same number of common PCs. For adaptive methods, the lack of a definition for the distribution of a curve may also arise some difficulties.

Ongoing and future work on our cluster analysis involves developing a clustering method, which is sensitive, at the same time, to both magnitude and shape heterogeneities of functional data.

References

[1] Arena, M. (2008). Does insurance market activity promote economic growth? A cross-country study for industrialized and developing countries. *Journal of Risk and Insurance, 75*(4), 921–946.

[2] Athanasiadis, S. & Mrkvička, T. (2018). European Insurance Market Analysis: A Multivariate Clustering approach. *Proceedings of the 12th International Scientific Conference INPROFORUM on Innovations, Enterprises, Regions and Management*, 328–334.

[3] Bouveyron, C., Côme, E. & Jacques, J. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics, 9*(4):1726–1760.

[4] Burden, R. L. & Faires, J. D. (2011). *Numerical analysis,* (9th edn). Boston, MA: Brooks/Cole, Cengage Learning.

[5] Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. (2014). Nbclust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software, 61*(6), 1–36.

[6] Dai, W., Mrkvicka, T., Sun, Y. & Genton, M. (2018). Functional Outlier Detection and Taxonomy by Sequential Transformations, *arXiv preprint arXiv: 1808.05414v1*.

[7] Delaigle, A. & Hall, P. (2010). Defining probability density for a distribution of random functions. *Ann Stat*., 38, 1171–1193.

[8] Dragos, S. L. (2014). Life and non-life insurance demand: The different effects of influence factors in emerging countries from Europe and Asia. *Economic Research-Ekonomska Istraživanja, 27*(1), 169–180.

[9] Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis.* New York: Springer.

[10] Galeano, P. & Pena, D. (2000). Multivariate Analysis in Vector Time Series. *Resenhas do Instituto de Matematica e Estatıstica da Universidade de Sao Paulo, 4*(4), 383–403.

[11] Hardle, W. & Simar, L. (2003). *Applied Multivariate Statistical Analysis*. Verlag Berlin Heidelberg: Springer.

[12] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6-7), 417–441; 498–520.

[13] Ieva, F., Paganoni, A. M., Pigoli, D., & Vitelli, V. (2012). Multivariate functional clustering for the morphological analysis of ECG Curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 62*, 401-418.

[14] Jacques, J. & Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification, 8*(3), 231–255.

[15] Kaufman, L. & Rousseeuw, P.J. (1990). *Partitioning around medoids (program pam). Finding groups in data: an introduction to cluster analysis*. Hoboken, NJ, USA: John Willey & Sons, Inc.

[16] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297.

[17] McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.

[18] Millo, G. & Carmeci, G. (2015). A sub-regional panel data analysis of life insurance consumption in Italy. *Journal of Risk and Insurance, 82*(2), 317-340.

[19] Montero, P. & Vilar, J.A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software*, *62*(1), 1–43.

[20] Nicol, F. (2013). Functional principal component analysis of aircraft trajectories. *Proceedings of the 2nd International Conference on Interdisciplinary Science for Innovative Air Traffic Management.*

[21] Olayungbo, D.O. & Akinlo, A.E. (2016). Insurance penetration and economic growth in Africa Dynamic effects analysis using Bayesian TVP VAR approach. *Cogent Economics & Finance*, *5*(1), 1–19.

[22] Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis,* (2nd edn.). New York: Springer.

[23] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461–464.

[24] Swiss Re Sigma Database. (2016). Zurich: Swiss Re Economic Research and Consulting.

Relative Income in Durable and Non-Durable Consumption Function: Cross-Sectional Study

Ondřej Badura[1]

Abstract. Since the life cycle hypothesis the consumption theory has been understood as the theory of utility streaming from consumption of goods and services at the time of its purchase. In this sense the consumption of durable goods is actually not a consumption, but rather a kind of investment. Although this statement seems to be clear and logic, for the sake of simplicity many empirical studies ignore this fact and assume the same structural patterns for both durable and non-durable consumption function. But can we actually simplify household consumption behaviour that much at the first place? Does this equivalence hold also for the interdependent concept of consumption and utility?

The aim of this work is to make a little contribution in answering these questions by estimating the consumption functions of durables and non-durables separately focusing on the influence of relative income. The regression analysis of the cross-sectional dataset of Czech households reveals that while relative income is the key determinant of non-durable consumption function, its influence is strongly insignificant in the durable consumption behaviour. This finding implies that interdependent nature of consumption truly holds only for the situations where the utility is gained at approximately the same time as the related consumption.

Keywords: consumption, durables, non-durables, relative income, cross-sectional regression.

JEL Classification: D12, D14
AMS Classification: 62P20

1 Introduction

It is not necessary to emphasize what is obvious: a consumption is the ultimate goal of any economic effort. So the study of the determinants of consumption behaviour is, in the same line, the investigation into the very nature of economics itself. There has been a great progress in the recognition of consumption behaviour since Keynes [15] first presented what we now call a consumption function. The main contribution can further be seen especially in the concept of life cycle – permanent income hypothesis (Modigliani and Brumberg [16]; Friedman [13]) later amended by the aspect of rational expectation by Hall [14] and his random walk model. Significant enrichments of consumption theory were later achieved in the concept of liquidity constrains and myopia (Flavin [11]), models combining proportional relationship of consumption and income with intertemporal preferences (Campbell and Mankiw [6]) or the precautionary motive and buffer-stock saving by Carroll [7]. Among recent studies investigating determinants of consumption behaviour we can mention for example Berger et al. [3], dealing with the influence of house prices or Aizenman, Cheung and Ito [1] who focus on the interest rate channel.

Although it is true that since the life cycle hypothesis the consumption theory has been understood as the theory of utility streaming from consumption of goods and services at approximately the time of its purchase, for the sake of simplicity many empirical studies ignore this fact and assume the same structural patterns for both durable and non-durable consumption function. But can we actually simplify household consumption behaviour that much at the first place? As indicated by studies dealing with this issue (Browning [5]; Fernández-Villaverde and Krueger [10]; Bertola, Guiso and Pistaferri [4]) this kind of simplification could easily lead to many premature and wrong conclusions.

Does the same apply also for the implications of interdependent concept of consumption and utility? Is the relationship based on relative standings or comparison (Duesenberry [9]) different in the durable and non-durable consumption behaviour? To make a little contribution in answering this question is the primary goal of this paper. The cross-sectional regression analysis of this work that is based on the large sample of Czech households is particularly focused on the relationship between the position of the given household in the income dis-

[1] VŠB - Technical University of Ostrava, Department of Economics, Sokolská třída 33, 702 00 Ostrava, Czech Republic, ondrej.badura@vsb.cz

tribution (approximated by relative income) and its average propensity to consume in the durable and non-durable consumption functions.

The structure of the paper is as follows. In the second section we introduce the data, the model and the method of estimation. Section 3 puts forward the result and briefly comments on them and in the final section 4 we conclude.

2 Data and model

To investigate the relative income effect in the durable and non-durable consumption on the example of the Czech households we use a cross-sectional regression analysis. Before describing its technical properties we need to introduce the input dataset first.

2.1 Data

As mentioned above we follow the consumption function of individual units, particularly we focus on the household-level behaviour. There is no other source of data for budgetary behaviour of Czech households large enough except for Household Budget Survey (HBS) run by Czech Statistical Office. We use a cross-sectional dataset, particularly the HBS wave from 2015 with the total number of 2929 observations.

There are characteristics of households that serve as initial variables entering the subsequent analysis, they are:

APC – average propensity to consume,
YR – relative income,
group – economic status of the head of the household:
1) employees with lower education (reference category),
2) self-employed,
3) employees with higher education,
4) households of inactive people with EA members,
5) retired (non-working) without EA members,
6) unemployed,
7) other households without EA members,

gen – gender of the head of the household:
1) man (reference category),
2) woman,

mort – mortgage:
1) no (reference category),
2) yes,

source – main source of income:
1) from work and equity (reference category),
2) social,

age – age of the head of the household.

The response variable - the average propensity to consume was derived according to the fundamental formula:

$$APC_i = \frac{C_i}{Y_i} \tag{1}$$

where C_i is consumption expenditures per month (per capita) of household i and Y_i is its net pecuniary income per month (per capita). While the value of Y_i in the Equation 1 does not change in modelling of non-durable or durable consumption, the variable C_i contains all consumption expenditures in the case of total consumption function, it contains only non-durables and services in the case of non-durable consumption function and it contains only semi-durables and durables in the case of durable consumption function. We further mark the average propensity to consume for of total consumption APC_total, for the non-durable consumption APC_non and for the durables APC_dur.

The relative income variable in the sense of Duesenberry [9] is particularly defined according to the formula introduced by Badura [2]:

$$YR_i = \frac{Y_i}{\frac{\sum_i^N Y_i w_i}{\sum_i^N w_i}} \tag{2}$$

where w_i stands for number of household members. In other words, Equation 2 says that relative income for each household is given by dividing its income by the weighted average of income in the society (of size *N*) that the household live in. As we can see, the weights in our case were set as the household size, which should, as Badura [2] points out, the best represent the influence of given household in the society.

After the first investigation of the data, 2 observations had to be dropped out as extreme outliers (their APC_total were 58.2 and 8.6 respectively). Therefore, the dataset entering the subsequent analysis contains 2927 observations.

2.2 Model

To achieve the goal of the paper that is investigating the relative income effect in non-durable and durable consumption function we introduce 3 models. The first one we call TOTAL and it represents the consumer behaviour based on the total consumption. It serves as a baseline model of the subsequent analysis and it is meant especially for the comparison. The structure of this model is as follows:

$$APC\_total_i = \beta_0 + \beta_1 YR_i + \sum_{j=2}^{7} \alpha_j group_{(j)i} + \beta_2 gen_{(2)i} + \beta_3 mort_{(2)i} + \beta_4 source_{(2)i} + \beta_5 age_i + \beta_6 age_i^2 + u_i \quad (3)$$

where average propensity to consume (for total consumption) of household *i* is explained by its relative income YR_i, set of dummy variables representing the economic status of the head of the household ($group_{(2)i}$ to $group_{(7)i}$), another dummy variable for gender of the head of the household $gen_{(2)i}$ (particularly taken value of 1 when the person is woman), dummy for having a mortgage $mort_{(2)i}$, another dummy taken value of 1 when social income is the main source of income $source_{(2)i}$ and age of the head which is included also in its squared form to reflect the U-shaped relationship, traditionally expected in the age-consumption relationship. There are set of estimated coefficients including the drift β_0, parameters $\beta_1 - \beta_6$ and $\alpha_2 - \alpha_7$. The last term in the equation u_i stands for the random error that is an influence of all other variables affecting APC not included into the model.

The other two estimated models are basically the same as the one described by the Equation 3 except for the response variable that is newly calculated only for the specific part of consumption. Particularly the NON-DURABLE model represents the consumption behaviour based on non-durables and it is given by a formula:

$$APC\_non_i = \beta_0 + \beta_1 YR_i + \sum_{j=2}^{7} \alpha_j group_{(j)i} + \beta_2 gen_{(2)i} + \beta_3 mort_{(2)i} + \beta_4 source_{(2)i} + \beta_5 age_i + \beta_6 age_i^2 + u_i \quad (4)$$

The model referred as a DURABLE then stands for specific characteristic of consumption function based on durables and it takes the following form:

$$APC\_dur_i = \beta_0 + \beta_1 YR_i + \sum_{j=2}^{7} \alpha_j group_{(j)i} + \beta_2 gen_{(2)i} + \beta_3 mort_{(2)i} + \beta_4 source_{(2)i} + \beta_5 age_i + \beta_6 age_i^2 + u_i \quad (5)$$

Since we use a cross-sectional dataset we don't have to use any special econometric method and the simple OLS can be used for all three models. Considering the goal of this work we make an crucial assumption of a negative value of the parameter β_1, because in line with the relative income hypothesis households standing higher in the income distribution should be able to safe higher fraction of their income and thus to relatively consume less, as demonstrated for example by Frank [12], Clark, Frijters and Shields [8] or Badura [2].

3 Results

Before we present the main results of the estimated models described above, let's first have a brief look at Figures 1-3 that represent a preliminary visual assessment of investigated relationship. We can easily identify a negative dependency between relative income and average propensity to consume out of total consumption (Figure 1), which is very similar to the visual assessment for the same relationship in the non-durable consumption behavior (Figure 2). On the other hand, Figure 3 reveals that the same pattern is probably not present in the consumption function for durables, where a linear dependency between YR and APC_dur is, at least, put in question.

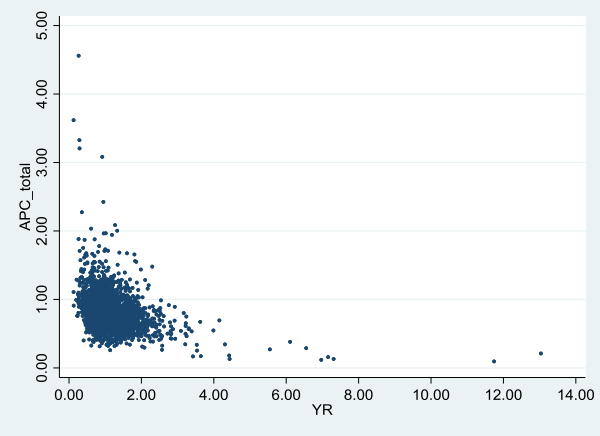


Figure 1 Visual assessment of YR and APC_total relationship

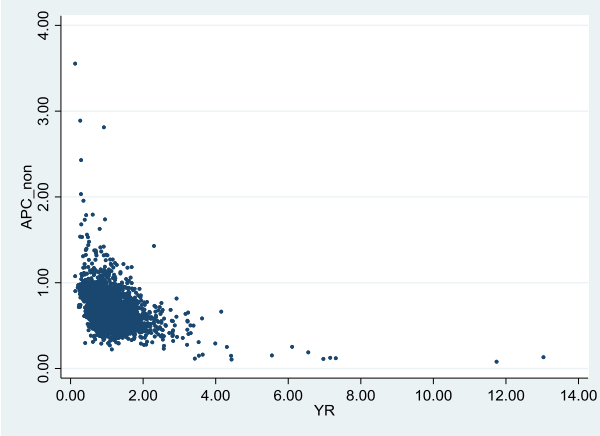


Figure 1 Visual assessment of YR and APC_total relationship

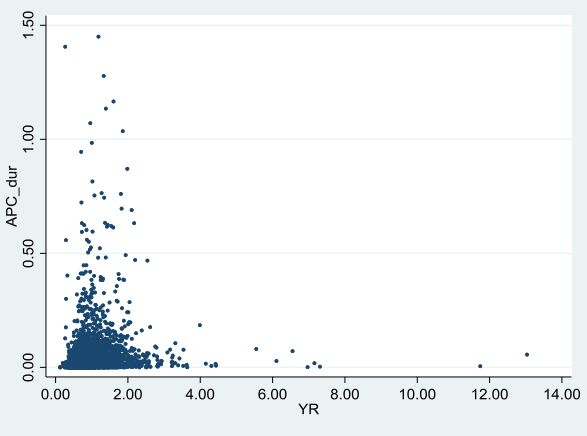


Figure 1 Visual assessment of YR and APC_total relationship

The coefficient estimations for all three models are summarized in the Table 1, where numbers in parenthesis are given p-values. Let's note first that in all cases Breusch-Pagan test indicates significant heteroscedasticity (with the test statistics of 652; 646 and 543 respectively) so in order to keep all subsequent t-tests and f-tests valid we had to estimate the coefficients with robust standard errors.

	TOTAL	**NON-DURABLE**	**DURABLE**
relative income	-0.105 (0.000)	-0.106 (0.000)	0.001 (0.846)
economic status of the head of the household:			
2) self-employed	0.098 (0.000)	0.085 (0.000)	0.013 (0.070)
3) employees with higher education	0.014 (0.138)	-0.005 (0.471)	0.02 (0.001)
4) households of inactive people with EA members	-0.03 (0.105)	-0.006 (0.689)	-0.024 (0.008)
5) retired (non-working) without EA members	0.044 (0.028)	0.054 (0.001)	-0.01 (0.365)
6) unemployed	0.105 (0.011)	0.137 (0.000)	-0.033 (0.052)
7) other households without EA members	0.122 (0.066)	0.103 (0.029)	0.019 (0.563)
gender (woman)	0.061 (0.000)	0.081 (0.000)	-0.02 (0.000)
mortgage	-0.08 (0.000)	-0.071 (0.000)	-0.009 (0.088)
main source of income (social)	0.109 (0.000)	0.056 (0.000)	0.052 (0.000)
age	0.01 (0.000)	0.007 (0.000)	0.003 (0.005)
age^2	-0.0001 (0.000)	-0.0001 (0.000)	-0.00004 (0.000)
R^2	0.194 (0.000)	0.268 (0.000)	0.064 (0.000)

Table 1 Estimation results

We can start the commentary on the Table 1 by saying that despite relatively low values of coefficients of determination all estimated models showed up to be statistically significant. While the low value of R^2, especially for the DURABLE model, indicates that many aspects influencing the response variable (APC) have still remained hidden within the error term. On the other hand, a high value of R^2 is only rarely seen in the cross-sectional analysis. Moreover, we are not interested in explanation of all structural patterns of given consumption behaviour, but only in the impact of one specific variable. With this in mind we can mark the R^2 for the TOTAL and NON-DURABLE models as sufficient enough, while its value for the DURABLE model may still be marked as suspiciously too low.

When commenting on the individual estimates let's first focus on the variable of our main interest – relative income. According to the result of our baseline model (TOTAL) it truly showed up to be significant determinant of household consumption behaviour. In line with the assumption we have made the value of its coefficient is negative and indicates that rise in relative standings in the income distributions by 1 (weighted average of income) also lowers the average propensity to consume (out of total consumption) of given household by something over 10 percentage points. What was first indicated by the previous visual assessment is now confirmed also by the estimation results that is the stability of APC and YR relationship in total consumption and non-durable consumer behaviour. As the value of the coefficient between TOTAL and NON-DURABLE models practically doesn't change we can conclude that relative income effect is virtually the same in both the total and the non-durable consumption functions.

A completely different picture of the same relationship is given by the DURABLE model. The coefficient for relative income variable is positive and, moreover, strongly insignificant. Not only that this finding underlines the principal differences between durable and non-durable spending, it also confirms that relative income effect is, in fact, present only in the situations where the utility is gained at the same time as is the given consumption.

Let's make also a brief note to the remaining results. When comparing the results for TOTAL and NON-DURABLE models we can see that similarly as in the case of relative income also most of other coefficients and their p-values don't significantly change. This fact indicates that not only relative income effect, but also the other patterns of consumer behaviour are very similar both in the total and non-durable consumption. While the influence of some factors remains quite similar also for the durable consumption behaviour – for example having the social main source of income or the age of the head of the household, other factors like gender or being unemployed have exactly the opposite impact. This finding only underlines the structural differences between patterns in spending for durables and non-durables.

4 Conclusion

The primary motivation of this work was to investigate the relative income effect in the non-durable and durable consumption functions. Our cross-sectional analysis of Czech households based on the Household Budget Survey data from the 2015 wave revealed relatively strong similarities in total and non-durable consumption behaviour and quite significant differences compared to the behaviour in spending for durables. As for the effect of relative income, it is no doubt that relative position in income distribution is one of the major factors affecting the consumption spending for non-durables, but on the other hand it is irrelevant in term of spending for durables. This finding only underlines the necessity of distinction between the consumption function for non-durables (and services) a durables.

In the same line, results of our analysis fully support the "keeping up with the Joneses" hypothesis (at least for the case of Czech households) implying that lower income households tend on average to spend more fraction of their total income for consumption just to catch up with others. The important conclusion of this work is that this effect is present only in the situations where the utility is gained at approximately the same time as is the related consumption that is the consumption of non-durables and services. When the stream of utility from consumption is gained long after the time of its purchase, relative income is no longer a significant determinant affecting the distribution of income into consumption and saving and the given spending behaviour is rather subject to investment patterns, where relative income effect is, as far as we know, irrelevant.

However these results are still the preliminary ones. It is necessary to keep in mind that this analysis was run only for the cross-sectional consumption function. Consumption behaviour in the time may imply a little different conclusions. Long story short, a lot of further analysis and robustness check will have to be made before it is possible to unambiguously confirm the principles stated above. Moreover relatively low values of coefficient of determination (especially for the durable consumption model) indicates that many influential variables have still remained hidden and so to find them a thus to rise the share of the variability explained may be one of the major motivation for extension of this work.

Acknowledgements

This paper was financially supported within the VŠB–Technical University SGS grant project No. SP2019/119 (Econometric Modeling of Fiscal, Monetary and Financial Aspects of the European Economies).

References

[1] Aizenman, J., Cheung, Y. and Ito, H. (2016). The Interest Rate Effect on Private Saving: Alternative Perspectives. *NBER Working Paper No. 22872*.
[2] Badura, O. (2018). Vliv relativního příjmu na sklon ke spotřebě – případ České republiky. *Politická ekonomie*, 66(4), 430-449.
[3] Berger, D. et al. (2018). House Prices and Consumer Spending. *Review of Economic Studies*, 85(3), 1502-1542.
[4] Bertola, G., Guiso, L. and Pistaferri, L. (2005). Uncertainty and consumer durables adjustment. *Review of Economic Studies*, 72(4), 973-1007.
[5] Browning, M. (1989). The Intertemporal Allocation of Expenditure on Non-Durables, Services, and Durables. *The Canadian Journal of Economics*, 22(1), 22-36.
[6] Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence. *NBER Macroeconomics Annual*.
[7] Carroll, Ch. D. (1997). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. *Quarterly Journal of Economics*, 107(1), 1-56.
[8] Clark, A. E., Frijters, P. and Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, 46(1), 95-144.
[9] Duesenberry, J. S. (1949). *Income, Saving, and the Theory of Consumer Behavior*. Cambridge: Harvard University Press.
[10] Fernández-Villaverde, J. and Krueger, D. (2011). Consumption and Saving over the Life Cycle: How Important are Consumer Durables? *Macroeconomic Dynamics*, 15(5), 725-770.
[11] Flavin, M. A. (1985). Excess Sensitivity of Consumption to Current Income: Liquidity Constraints or Myopia? *Canadian Journal of Economics*, 18(1), 117-136.
[12] Frank, R. H. (1985). The Demand for Unobservable and Other Nonpositional Goods. *The American Economic Review*, 75(1), 101-116.
[13] Friedman, M. (1957). *A Theory of the Consumption Function*. New York: Princeton University Press.
[14] Hall, R. E. (1978). Stochastic Implication of the Life Cycle – Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*, 86(6), 971-987.
[15] Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.
[16] Modigliani, F. and Brumberg, R. (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In K. Kurihara (Eds.), *Post-Keynesian Economics*. New Brunswick: Rutgers University Press.

Possible Statistical Comparison of Two Time Series

Tomáš Ťoupal[1]

Abstract. The most often discussed problem is the comparison of two time series (for example financial or non-financial time series). The proposed approach can be assessed on the base of knowledge of the set of past values and the assumption that there is no significant change in the used probability model. The presented paper is motivated by the approach of positive and quadrant dependence. There is a "power of matching" between two time series, which can be measured (or estimated) in many ways. The problem may occur with their non-stationarity. One solution to this problem can be the quantification of preserving (or not preserving) probability of a monotone relationship. It means the probability that the values of the first time series are increasing and the values of the second time series are also increasing and similarly for consistent decreases. One measure for this quantification is proposed here followed by application on real data sets (Prague Stock Exchange) to estimate the price of one asset depending on the price of another asset.

Keywords: Statistical comparison, Markov model, Prague Stock Exchange, Coefficient of concordance, dependence, time series.

JEL Classification: C51, C13
AMS Classification: 91B82, 62-07

1 Introduction

The problem of comparing two time series is still a popular topic of discussion in a wide range of interest. Here are, for example, financial or non-financial time series, such as unemployment rate, GDP or stock markets. A common question (but not very often discussed) is whether two samples are similar in the case of chronological order. The desired response can be obtained from a set of past values (relative to the chronological order) and on the assumption that there is no significant change in the probability behavior model of such a sample.

The main idea of this proposed approach is therefore based on the main assumption of the degree of "power (or matching)" of keeping or not keeping the same tendencies (monotony, increase, decrease) in two time series. This means that the main aim of this work is to find or estimate a possible "degree" of matching on a real data set, where these results can be used for trading on the Prague Stock Exchange.

1.1 Used terms

The proposed approach is motivated by the concept of positive and quadrant dependence (e.g. [1] and others). The "power of matching" between two time series can (certainly) be measured in many ways. One measure can present probability of monotonous relationship. This can be explained by the probability that the values of both time series increase or decrease (this can also be used for mismatched behavior).

Only samples whose values are continuous random variables (this model does not affect reality) are used here. Therefore, the following table shows possible cases that may occur during the proposed approach.

Sample A	increases	increases	decreases	decreases
Sample B	increases	decreases	increases	decreases
Status code	1,1	1,0	0,1	0,0
Numeric code	3	2	1	0

Table 1 Proposed states

This approach can be extended to other system states (for example no movement). It should be noted that the increase or decrease must be related to some non-empty interval or to compare the value at one "time point" to the value at another "time point". For this reason, used samples are limited to time series with discrete time axis and for comparison at two time points. The first step is to determine a time shift "appropriately". This time shift, denoted as τ, gives possible (and predicted) states listed in the following table.

[1] European Centre of Excellence NTIS – New Technologies for Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 301 00 Plzeň, Czech Republic, e-mail: ttoupal@kma.zcu.cz.

Sample A	$x_A(t) > x_A(t-\tau)$	$x_A(t) > x_A(t-\tau)$	$x_A(t) < x_A(t-\tau)$	$x_A(t) < x_A(t-\tau)$
Sample B	$x_B(t) > x_B(t-\tau)$	$x_B(t) < x_B(t-\tau)$	$x_B(t) > x_B(t-\tau)$	$x_B(t) < x_B(t-\tau)$
Status code	1,1	1,0	0,1	0,0
Numeric code	3	2	1	0

Table 2 States of one comparison between two samples

Now, the question is which time shift from all shifts is correct. Discussing this issue is at the end of this article. This step causes the state to be assigned to each pair of time series (for given τ and at each time t). This is particularly interesting for selecting one time shift τ at each time point. It is also important to get a certain "degree of consistency (degree of matching)" denoted as the concordance rate. The concordance rate in this paper is the probability of the same movement. This means the probability of two states $(1,1)$ and $(0,0)$ at any time. There is worked with a pair of time series, for which this probability does not depend (functionally) on time t (but may and will depend on time shift τ). This relationship, which meets previous assumptions, is

$$\begin{aligned} concordance_r(x_A, x_B; \tau) &= P([1,1]|_\tau \cup [0,0]|_\tau) \\ &= P([1,1]|_\tau) + P([0,0]|_\tau) \\ &\equiv P([3]|_\tau) + P([0]|_\tau). \end{aligned} \tag{1.1}$$

And the following relationships are valid

1. $concordance_r(x_A, x_B; \tau) = concordance_r(x_B, x_A; \tau)$,
2. $0 \leq concordance_r(x_A, x_B; \tau) \leq 1$,
3. $x_B(t) = h(x_A(t)) \Rightarrow concordance_r(x_A, x_B; \tau) = 0, h(x)\ is\ decreasing$,
4. $concordance_r(x, x_*; \tau) = 1$, *where* x_* *is a copy of* x.

(1.2)

In analogy to Kendall's tau [2], used coefficient of concordance can be rearranged as a difference between probability match and probability mismatch. This rearrangement is in the following formula

$$\begin{aligned} concordance_c(x_A, x_B; \tau) &= P([1,1]|_\tau \cup [0,0]|_\tau) - P([1,0]|_\tau \cup [0,1]|_\tau) \\ &= 2concordance_r(x_A, x_B; \tau) - 1. \end{aligned} \tag{1.3}$$

The relationships (1.2) are also rearranged as

1. $-1 \leq concordance_c(x_A, x_B; \tau) \leq +1$,
2. $concordance_c(x_A, x_B; \tau) = 0$, *if decreases or increases in both series are independent*
3. $concordance_c(x_A, x_B; \tau) = -1$, *if increases in one series implies decreases in second series,*
4. $concordance_c(x_A, x_B; \tau) = +1$, *if increases in one series implies increases in second series.*

(1.4)

This approach brings nothing new. If the coefficient is negative, the probability of mismatch in the behavior of time series outweighs the probability of the same behavior. If the coefficient is positive, the probability of the same behavior outweighs the probability of mismatch behavior. If it is not possible to make a "significant" decision from the behavior of one time series to the behavior of the second time series, it is zero or almost zero.

1.2 Statistical view

A pair of time series in any state implements state events at a given time, with a selected time shift. These events are equally distributed. However, the assumption of observation independence will not be met for most real (pairs) time series (this is necessary if we want to use a random sample, i.i.d.). A pair of time series will have more or less amount of memory (= probability of dependence over time). It is thus possible to use the Markov chain model (a model of time dependence specific to common pairs of time series without further functional relationships, homogeneous and regular [2]). This approach replaces the assumption of independence of the observed states by assuming the independence of the observed transitions from one state to another.

For the classical Markov chain model holds

$$p^T(t+1) = p^T(t)P, \tag{1.5}$$

where

$p(t)$ is a probability vector of the occurrence of Markov chain in a given state and

P is a probability matrix of transition from state to state, obviously $dim\ p(t) = k$ and $dim\ P = kx\ k$, in this case is $k = 4$.

The homogeneous Markov chain model does not depend on the number of steps, i.e. on time (it does not age). The required probabilities can be estimated by selected statistical verification, which can be obtained by classical relative frequencies

$$\hat{P}(i,j) = \frac{n(i,j)}{n(i)}, \tag{1.6}$$

where $\hat{P}(i,j)$ is a probability of transition to state j from state i, $n(i,j)$ is a number of observed transitions from state i to state j and $n(i) = \sum_{l=1}^{k} n(i,l)$. $\hat{P}(i,j)$ is not the ratio of random variable of the number of observations and the fixed range of random sample. In this case, it is the ratio of two random variables. It makes the required estimate difficult. However, such an estimate is a consistent[2] estimate of $P(i,j)$, which is sufficient for our purposes.

Another problem can be in the properties of considered matrix P. Here, it is $P(i,j) > 0$, which means that the probability of being in a particular state "in one step" and that it can go into any other step or remain is non-zero. This is not guaranteed by the above-mentioned frequency approach (especially with a small number of observations called a zero frequency problem). Because of this situation, the Bayesian approach is used, assuming a uniform a priori distribution (e.g. [3]):

$$\hat{P}(i,j) = \frac{n(i,j)+1}{n(i)+k}; \quad i,j = 0,1,2,3. \tag{1.7}$$

It is a consistent estimate again and $k = 4$ in this case.

2 Real data application

2.1 Used data sets

It is necessary to have a “sufficient history” of data sets for real application. For this reason, selected data sets from the Prague Stock Exchange are used here. There are more usable stock titles with possible “longest history”. The following stock titles are selected because of their history since 1998. For further calculations, the stock data sets of the following companies are used:

- CEZ Group (as CEZ),
- Commercial Bank (as KB),
- O2 Czech Republic (as O2),
- Philip Morris International in the Czech Republic (as PM),
- Unipetrol (as UNI).

2.2 Obtained results

For the sake of clarity, the following tables illustrate the relationships between the selected companies with the highest value of the coefficient of concordance and the lowest value of the coefficient of concordance. The first is the relationship between CEZ Group and Commercial Bank. This result reaches the highest value of the coefficient of concordance of selected companies. The second highest value of the coefficient of concordance is reached for CEZ Group and O2.

[2] Proof can be done with using the characteristics of multinomial distribution.

From state (row)		Estimation of transition matrix, to state (column)			
		0,0	***0,1***	***1,0***	***1,1***
		0	***1***	***2***	***3***
0,0	***0***	***0,700***	*0,118*	*0,116*	*0,066*
0,1	***1***	*0,220*	***0,566***	*0,039*	*0,174*
1,0	***2***	*0,227*	*0,042*	***0,558***	*0,172*
1,1	***3***	*0,069*	*0,111*	*0,125*	***0,695***

Here, the probabilities of staying in concordant states dominate and the same increase or decrease is dominant outside these states.

Stationary probabilities		
P (0,0) =	P (0) =	***0,343***
P (0,1) =	P (1) =	***0,185***
P (1,0) =	P (2) =	***0,187***
P (1,1) =	P (3) =	***0,285***
Suma =		*1,000*
Concordance rate		***62,84 %***
Concordance coefficient		***0,257***

In the table of stationary probabilities, the dominance of the probabilities of occurrence in concordant states is evident. The probability of consistent behavior is high here, about 63 %.

Mean recurrence (stay) time		
t (0,0) =	t (0) =	***3,33***
t (0,1) =	t (1) =	***2,31***
t (1,0) =	t (2) =	***2,26***
t (1,1) =	t (3) =	***3,28***

Estimates of mean recurrence time in states again prefer concordant states.

From state (row)		Matrix of mean transition times, to state (column)			
		0,0	***0,1***	***1,0***	***1,1***
		0	***1***	***2***	***3***
0,0	***0***	***2,91***	*9,95*	*9,83*	*9,46*
0,1	***1***	*6,01*	***5,41***	*11,15*	*7,82*
1,0	***2***	*5,91*	*11,35*	***5,36***	*7,87*
1,1	***3***	*7,88*	*10,18*	*9,58*	***3,51***

Time series of both companies tend to concordant states. There is a faster movement to a concordant state rather than a non-concordant state.

The obtained result can be interpreted to mean that the time series of share price of both companies (CEZ Group and Commercial Bank) are in a "stronger" relationship.

Table 3 Comparison between CEZ Group and Commercial Bank

On the other hand, the lowest value of concordance rate is with PM company. One result obtained for PM and Unipetrol is shown in Table 4 below.

From state (row)		Estimation of transition matrix, to state (column)			
		0,0	***0,1***	***1,0***	***1,1***
		0	***1***	***2***	***3***
0,0	***0***	***0,645***	*0,150*	*0,154*	*0,050*
0,1	***1***	*0,215*	***0,564***	*0,041*	*0,180*
1,0	***2***	*0,201*	*0,049*	***0,581***	*0,169*
1,1	***3***	*0,082*	*0,151*	*0,137*	***0,630***

Probability of staying in the same position dominates.

Stationary probabilities		
P (0,0) =	P (0) =	***0,314***
P (0,1) =	P (1) =	***0,219***
P (1,0) =	P (2) =	***0,218***
P (1,1) =	P (3) =	***0,249***
Suma =		*1,000*
Concordance rate		***56,29%***
Concordance coefficient		***0,126***

The probability of occurrence in individual states is almost uniformly distributed. The probability of consistent behavior is not as high as before, it is about 56 %.

Mean recurrence (stay) time		
t (0,0) =	t (0) =	***2,82***
t (0,1) =	t (1) =	***2,29***
t (1,0) =	t (2) =	***2,39***
t (1,1) =	t (3) =	***2,71***

In terms of mean recurrence time in states, individual states are almost the same.

From state (row)		Matrix of mean transition times, to state (column)			
		0,0	***0,1***	***1,0***	***1,1***
		0	***1***	***2***	***3***
0,0	***0***	***3,19***	*8,09*	*8,21*	*9,49*
0,1	***1***	*5,89*	***4,57***	*9,88*	*7,71*
1,0	***2***	*6,03*	*9,51*	***4,58***	*7,84*
1,1	***3***	*7,34*	*8,02*	*8,55*	***4,01***

Both time series tend to have the same states and the deviation from the state causes faster movement (return) to the same state than to another state.

Here is one of the lowest levels of concordance coefficient from the selected set of real data. From the values of one time series, it is not possible to make "correct" conclusions about the price of another time series (in terms of unquantified changes, only increases and decreases).

Table 4 Comparison between Philip Morris and Unipetrol.

The graphical presentation of obtained results is in the following figures.

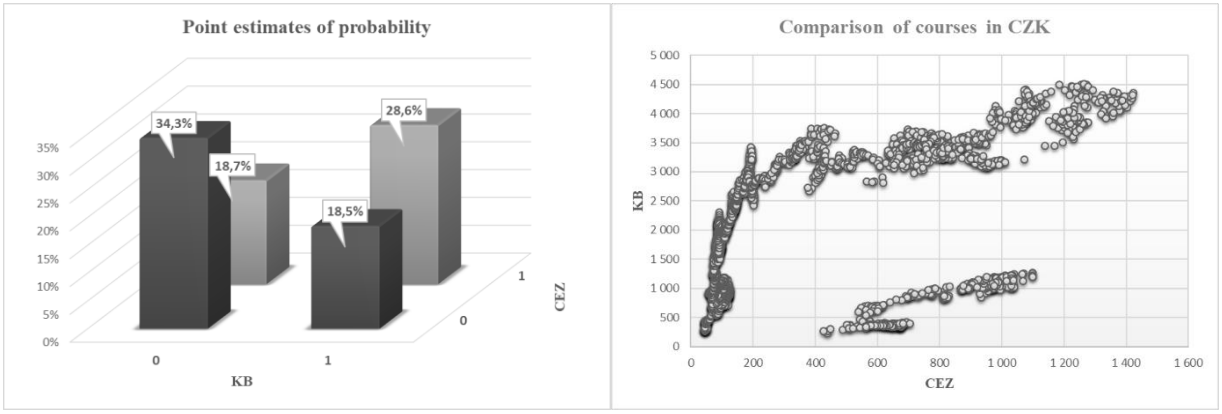


Figure 1 Comparison between CEZ Group and Commercial Bank

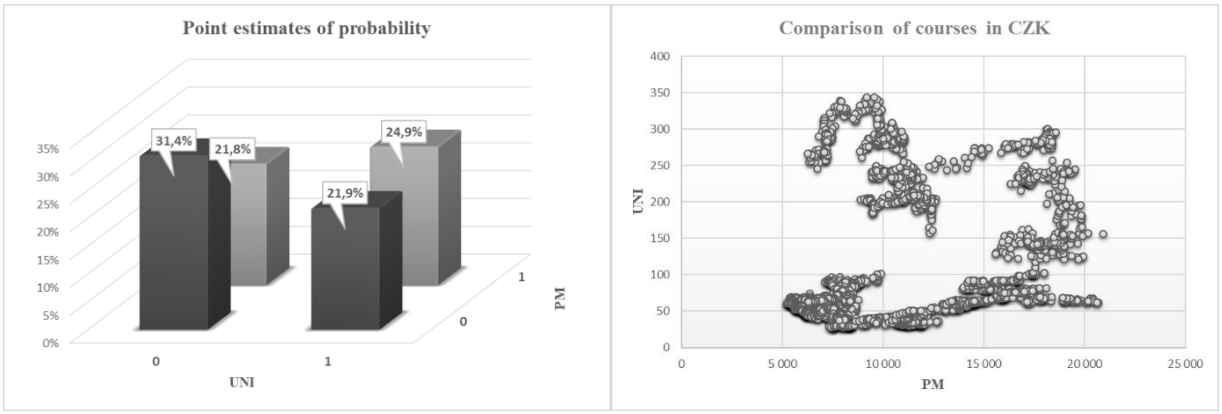


Figure 2 Comparison between Unipetrol and PM

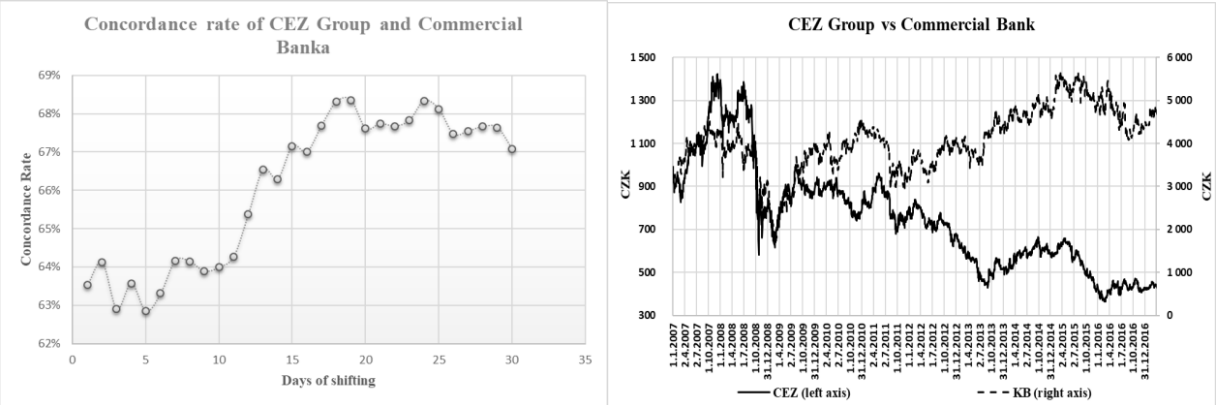


Figure 3 Influence of time shift and prices of CEZ Group and Commercial Bank

As shown in the graph above, the obtained results are different for different time shifts. The difference interval is about 6 %. The problem (and possible direction of research) is universal choice of time shift because plotted curve is different for each time series pair. With this information, a 5-day time shift is used for clarity based on observed simulations and for practical use (the obtained results are for the lowest risk and for the practical reason of one working or trading week). For the meaningfulness of these approaches, the maximum number of shifting days in simulations was limited to 30.

3 Conclusion

This paper deals with possible (not only) comparison of two non-stationary time series (pair comparison). The proposed approach (solution) to this problem is based on quantification of preserving (or not preserving) probability of monotone relationship. The published conclusions are for daily stock exchange prices of the Prague Stock Exchange:

Source	Count of Share	Currency	From the period	to
BCPP	5	CZK	5. 1. 1998	19. 4. 2018

Table 5 Source data set

The proposed concordance rate is practically (at a selected interval of 1 - 30 quoted days) described as a "probability match" independent of the "distance" of the data being compared. There are used two estimates of concordance rate and other statistics based on source data set from BCPP. The first estimate is with a high coefficient of concordance (it is for CEZ Group with others) and the second estimate is with a low coefficient of concordance (it is for PM with others). These obtained results may already be used for possible trading on the Prague Stock Exchange.

Acknowledgements

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

References

[1] Lehmann, E.L. (1966), *Some Concepts of Dependence*, [online] Ann. Math. Statist. Volume 37, Number 5, p. 1137-1153. Available at: <https://projecteuclid.org/euclid.aoms/1177699260> [Accessed 1 March 2018].

[2] Grinstead, CH.M. & Snell, J.L. (2003), *Introduction to Probability*, [online] Published by the American Mathematical Society. Available at: <https://math.dartmouth.edu/~prob/prob/prob.pdf> [Accessed 1 March 2018].

[3] Witten I. H. & Bell, T. C. (1991), *The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression*, [online] IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 37, NO. 4. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=87000> [Accessed 1 March 2018].

Dynamic AD-AS macroeconomic model of Mankiw type with generalized expectations

Ladislav Lukáš[1]

Abstract. The paper deals with formulation of discrete dynamic AD-AS macroeconomic model of Mankiw type. First, we discuss all macroeconomic structural equations in Mankiw setting. There are the following ones: the demand for goods and services, the Fisher equation expressing real interest rate, the Phillips curve expressing the inflation, adaptive expectation expressing expected inflation, and the monetary-policy rule expressing the nominal interest rate. Next, we focus our main attention to various general possibilities to handle expected inflation. One of the basic general expectations is formed as convex combinations of past inflations, which are discussed in detail together with numerical comparisons of results. We present the numerical implementation of dynamic AD-AS macroeconomic model of Mankiw type both with classical and generalized expectations in Mathematica.

Keywords: discrete dynamic AD-AS model, expected inflation, convex combination of past inflations, numerical implementation in Mathematica.

JEL classification: C51, E12
AMS classification: 39A12, 91B55

1 Introduction

The paper deals with formulation of discrete dynamic AD-AS macroeconomic model of Mankiw type enhanced with generalized expectation of inflation. For the theoretical background, we select three famous monographs of macroeconomics, [2], [3], and [4], in particular. We also amend just three jornal papers dealing with various aspects of macroeconomic modeling, [1], [5], and finally [6].

In [1], we can find an innovative methodology to build forecast of macroeconomic indicators that extends lag variables with some unstructured data in the form of financial news. In [5], the authors present a new comparative approach to macroeconomic modeling and policy analysis. Using this approach a model archive is built that includes many well-known empirically estimated models that can be used for quantitative analysis of monetary and fiscal stabilization policies. Finally, in [6], we can find interesting proposals how to teach and study macroeconomics with data related to the three main areas of macroeconomics, i.e. the short-run economy (economic fluctuations), the long-run economy (classical economic theory), and the very long-run economy (economic growth), too.

2 Macroeconomic model of Mankiw type

In this section we follow ideas and model construction given in [4], Ch. 14. The model is called *dynamic model of aggregate demand and aggregate supply* there. An interesting feture of the model is that it offers anothyer lens through which to view the business cycle and the effects of monetary and fiscal policy, respectively.

Compared with other popular macroeconomic models, the dynamic AD-AS model is closer to those studied by economists at the research frontier, which elevates its atractivity. Now, we introduce particular elements of the model with notations used.

Let Y_t denote national income in time period t, $\bar{Y}_t$ is the economy's natural level of output, r_t is the real interest rate, ϵ_t is a random demand shock, and $\alpha > 0$ and $\rho > 0$ are positive parameters given.

The demand for goods and services in time period t is given by Eq. (1)

$$Y_t = \bar{Y}_t - \alpha(r_t - \rho) + \epsilon_t. \tag{1}$$

The key feature of this equation is the negative relationship between the current real interest rate r_t and the demand for goods and services Y_t. It is well-kown that when the r_t increases, borrowing money becomes more expensive, and on the contrary, saving yields a greater reward. It causes, firms engage in fewer investment projects, and consumers save more and spend less. Both of these effects reduce the demand for goods and services.

[1]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

The parameter α expresses proportionality between the demand for goods and services and changes of real interest rate r_t being lessen by parameter ρ, respectively. From a mathematical perspective, ρ is just a given constant, but it has a useful economic interpretation. It corresponds to the real interest rate at which, in absence of any exogeneous shock ($\epsilon_t = 0$), the demand for goods and services equals the natural level of output, i.e. $Y_t = \bar{Y}_t$, which is simply evident substituting, $\epsilon_t = 0$ and $\rho = r_t$, into the Eq. (1).

The first term on the right-hand side of Eq. (1), $\bar{Y}_t$, implies that the demand for goods and services rises with the economy's natural level of output. In this general setting it can be time-dependent. However, in the most cases, one simply assumes it to be constant for every time period t. On the contrary, investigating long-run economy growth represented by exogenous increases in $\bar{Y}_t$, one needs a sequence of this quantity to be given in advance for all time periods t considered.

The last term in the demand equation (1), $\epsilon_t \in N(0, \sigma_\epsilon)$, is a random variable which represents exogeneous impacts in the demand assuming the variance $\sigma_\epsilon^2 > 0$ is given. The exogeneous quantity ϵ_t, also captures changes in fiscal policies that affect the demand of goods and services, in general.

The second component of the model is formation of real interest rate, r_t, using well-known Fisher equation, as follows

$$r_t = i_t - E_t\pi_{t+1}. \tag{2}$$

As usual, the real interest rate r_i is expressed as a difference between nominal interest rate i_t and the expected rate of future inflation $E_t\pi_{t+1}$. The variable r_t is more precisely known as *ex ante* real interest rate. The variables r_t and i_t are interest rates that prevail at time t and are known at t and, therefore, represent corresponding rates of return between periods t and $t + 1$.

The variable π_t denotes the current inflation rate, which is the percentage change in the price level between periods $t - 1$ and t, as it is well-known classical formulation. Similarly, π_{t+1} represents the inflation rate that will occur between periods t and $t + 1$, and obviously, it remains unknown at time period t. On the contraty, the expectation of future inflation $E_t\pi_{t+1}$ is known at period t. Besides *ex ante* real interest rate, given by $i_t - E_t\pi_{t+1}$, one also knows the *ex post* real interest rate, which is given by $i_t - \pi_{t+1}$, and will not be known until period $t + 1$, evidently.

The third component of the model is formation of inflation, π_t, using conventional Phillips curve augmented to include roles for expected inflation and exogeneous supply shocks, as follows

$$\pi_t = E_{t-1}\pi_t + \phi(Y_t - \bar{Y}_t) + \eta_t, \tag{3}$$

where $E_{t-1}\pi_t$ expresses previously expected inflation built at period $t - 1$, quantity $(Y_t - \bar{Y}_t)$ gives the deviation of output from its natural level, and finally, η_t is a random variable, $\eta_t \in N(0, \sigma_\eta)$, with the variance $\sigma_\eta^2 > 0$ given, representing a supply shock applied at t, which should include all exogeneous events that directly influence inflation.

The parameter $\phi > 0$ is given, and expresses proportional response of inflation upon current fluctuation of Y_t around its natural level $\bar{Y}_t$.

A specific feature of the Mankiw model is that it does not include unemployment explicitly, but it assumes, on the base of Okun's law, that short-run fluctuations in output and unemployment are strongly and negatively correlated. Hence, the unemployment fluctuates along with output, i.e. $(Y_t - \bar{Y}_t)$, but in the opposite direction.

The fourth component of the model is formation of expected inflation, $E_t\pi_{t+1}$. In reality, it is open for creativity, and we shall grasp it, as will be evident later on. However, the original version of Mankiw model proposes the simplest formula, which can be written as

$$E_t\pi_{t+1} = \pi_t. \tag{4}$$

This assumption about inflation expectations is admittedly crude. It means nothing else, but that expected inflation in period $t + 1$, $E_t\pi_{t+1}$, will be the same as recently observed inflation rate, π_t. It is sometimes called adaptive expectation, too, and it also simplifies solution of the model by reducing number of variables to be determined.

Finaly, the fifth component of the model is formation of nominal interest rate, i_t, which is a core of monetary policy, as follows

$$i_t = \pi_t + \rho + \theta_\pi(\pi_t - \tilde{\pi}_t) + \theta_Y(Y_t - \bar{Y}_t). \tag{5}$$

Recall that ρ, the constant in this equation, is the natural rate of interest, i.e. the real interest rate at which, in the absence of any shock, the $Y_t = \bar{Y}_t$, and π_t is inflation rate.

The Eq. (5), containts also two terms expressing fluctuations. First, the fluctuation $(\pi_t - \tilde{\pi}_t)$, of inflation rate π_t around the central bank's target for the inflation rate, denoted $\tilde{\pi}_t$. Second, the fluctuation of the output $(Y_t - \bar{Y}_t)$, which has been already discussed. These fluctuations are multiplied by monetary policy parameters θ_π and θ_Y, which are non-negative ones, in general. They indicate how much the central bank allows the interest rate target to respond to fluctuations in inflation and output, respectively.

Similar to $\bar{Y}_t$, the target inflation rate $\tilde{\pi}_t$, can be usually assumed to be constant, but one may accept to vary it over time periods in particular cases, too.

In general, the Eq. (5), tells us how how the target for nominal interest rate i_t, chosen by the central bank responds to macroeconomic conditions. To interpret this equation, it is best to focus not just on nthe nominal interest rate i_t, but also on the real interest rate r_t. We know, that the r_t, rather than i_t, influences the demand for goods and services. So, although the central bank sets a target for the nominal interest rate $\tilde{i}_t$, the bank's influence on the economy works through the real interest rate r_t, at all. By definition, $r_t = i_t - E_t\pi_{t+1}$, and using the Eq. (4), one simply gets relation $r_t = i_t - \pi_t$. Now, substituting it into Eq. (5), and assuming that inflation is at its target, i.e. $(\pi_t = \tilde{\pi}_t)$, and the output is at its natural level, i.e. $(Y_t = \bar{Y}_t)$, which cause vanishing the last two terms in this equation, evidently, we may conclude that $r_t = \rho$, i.e. the real interest rate equals the natural rate of interest.

Another specific feature of the Mankiw model is that it does not include money supply explicitly. Well, money supply is typically taken to be the policy instrument of the central bank, and the interest rate adjusted to bring money supply and money demand into equilibrium. The Mankiw model turns this logic on its head. The central bank is assumed to set a target for the nominal interest rate, $\tilde{i}_t$. It then adjust the money supply to whatever level is necessary to ensure that the equilibrium interest rate, which balances money supply and demand, hits the target.

The main advantage of using the interest rate, rather than the money supply, as the policy instrument in the dynamic AD-AS model is, that it is more realistic. Today, most central banks set a short-term target for the nominal interest rate. One has to keep in mind, though, that hitting that target $\tilde{i}_t$, requires adjustment in the money supply. When a central bank decides to change the interest rate, it is also committing itself to adjust the money supply accordingly.

*Note*1: In the model, the Eq. (1), represents a base for construction of aggregated demand (AD) of product. While, aggregated supply (AS) of product is yielded from Eq. (3), just by rearranging it into proper form, as follows

$$Y_t = \bar{Y}_t + \frac{1}{\phi}(\pi_t - E_{t-1}\pi_t - \eta_t). \tag{6}$$

The aggregated demand (AD) of product is constructed from Eq. (1), using Eq. (2) to substitute r_t, exploiting Eq. (4) to replace $E_t\pi_{t+1}$ by π_t, and finally, using Eq. (5) to substitute i_t, as follows

$$Y_t = \bar{Y}_t - \alpha(\theta_\pi(\pi_t - \tilde{\pi}_t) + \theta_Y(Y_t - \bar{Y}_t)) + \epsilon_t),$$

and finally

$$Y_t = \bar{Y}_t - \frac{1}{1 + \alpha\theta_Y}(\alpha\theta_\pi(\pi_t - \tilde{\pi}_t) - \epsilon_t). \tag{7}$$

*Note*2: One can simply express equations of AD and AS of product in popular form containing dynamic production gap, as follows

$$\begin{aligned} Y_t - \bar{Y}_t &= -\frac{1}{1 + \alpha\theta_Y}(\alpha\theta_\pi(\pi_t - \tilde{\pi}_t) - \epsilon_t), \\ Y_t - \bar{Y}_t &= \frac{1}{\phi}(\pi_t - E_{t-1}\pi_t - \eta_i). \end{aligned} \tag{8}$$

2.1 Formation of generalized expectation

Forecasting, or formation of expectations, is a fruitful field of activity. In the theory, one knows an approach called rational expectation, according to which people optimally use all available information when forecasting the future.

In [3], Ch. 5.6.1, there is a folowing scheme for "adaptive expectation", where expectation is built at time period t, and the expected value of a quantity ξ at future period $t + 1$ is denoted, $\hat{\xi}_{t+1}$

$$\hat{\xi}_{t+1} = \kappa \sum_{j=0}^{\infty} (1 - \kappa)^j \xi_{t-j}, \tag{9}$$

where $\kappa \in (0, 1)$ is constant given in order to maintain convergence of the summation. However, the formula (10) resembling discounting paradigm, is rather academic one, as it contains an infinite sequence of past values of quantity ξ.

As a special case of Eq. (10), we restrict the sequence of past values of quntity ξ upon two recent ones, ξ_t, ξ_{t-1}, only, and adopt the following convex combination to form get $\hat{\xi}_{t+1}$

$$\hat{\xi}_{t+1} = \psi(\xi_t, \xi_{t-1}) = \kappa\xi_t + (1-\kappa)\xi_{t-1}, \tag{10}$$

with two alternatives for κ

- deterministic case: $\kappa \in (0,1)$ constant given,
- stochastic case: $\kappa \in U(0,1)\,\&\,\kappa \geq \kappa_b,\ \kappa_b \in (0,1)$ giving a filtration lower bound, respectively.

2.2 Model of Mankiw type with generalized expectation of inflation

The model is composed form Eq-s. (1), (2), (3), and (5), of original Mankiw model, and replacing Eq. (4) by the Eq. (10), thus instead a simple expectation of inflation we use more general version of expectation.

$$\begin{aligned} Y_t &= \bar{Y}_t - \alpha(r_t - \rho) + \epsilon_t, \\ r_t &= i_t - E_t\pi_{t+1}, \\ \pi_t &= E_{t-1}\pi_t + \phi(Y_t - \bar{Y}_t) + \eta_t, \\ E_t\pi_{t+1} &= \psi(\pi_t, \pi_{t-1}) = \kappa\pi_t + (1-\kappa)\pi_{t-1}, \\ i_t &= \pi_t + \rho + \theta_\pi(\pi_t - \tilde{\pi}_t) + \theta_Y(Y_t - \bar{Y}_t). \end{aligned} \tag{11}$$

3 Numerical simulations

In this section, we present several numerical simlations with the model. All calculations were performed by our Mathematica notebook, and the core of solving procedure is built by recurrent calls of Mathematica function NSolve[ ] to calculate all variables of the model. A pseudo-random generator was initiated by command RandomSeed[12321], in all simulated cases.

In Fig. 1, we calculate development of product Y_t in time $t = 1, \ldots, T = 20$, and inflation π_t, real interest rate r_t, and nominal interest rate i_t, respectively, assuming a simple exogeneous shocks.

Data: $\bar{Y}_t = 100,\ \tilde{\pi}_t = 2,\ \alpha = 1,\ \rho = 2,\ \phi = 0.25,\ \theta_\pi = 0.5,\ \theta_Y = 0.5,\ E_{t-1}\pi_t|_{t=1} = 1.9,\ \epsilon_t = 0,$ except $\epsilon_2 = 1$, representing a unit exogeneous shock at $t = 2\ \eta_t = 0$.

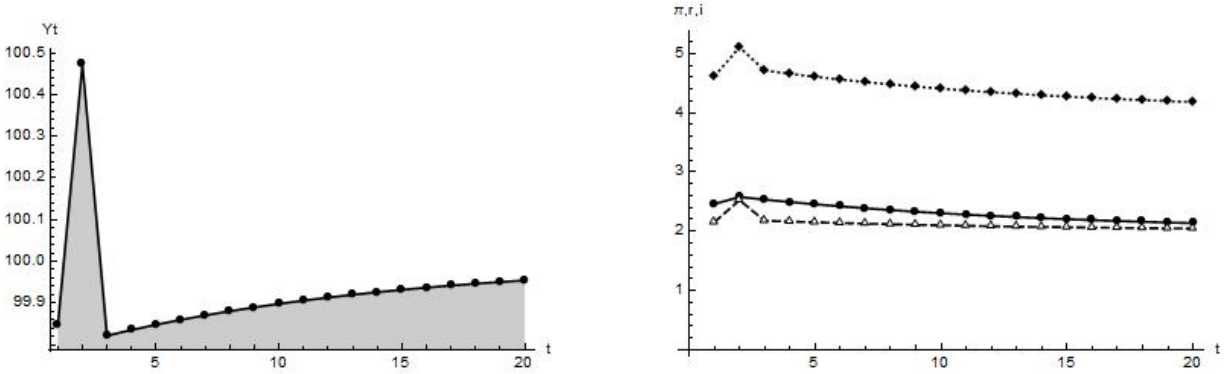
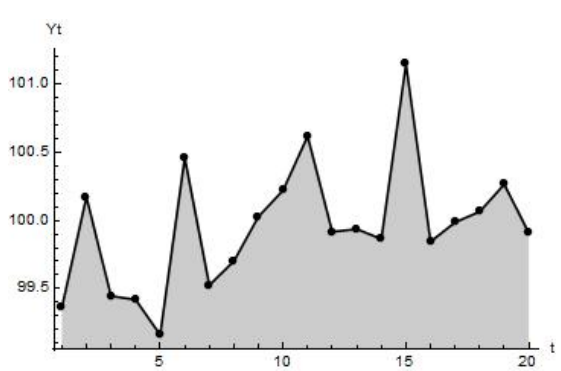


Figure 1: Left panel: Product Y_t, with shock $\epsilon_2 = 1$. Right panel: π_t (full line), r_t (dashed), i_t (dotted).

In Fig. 2, we use the same data except the exogeneous shock $\epsilon_t \in N(0, \sigma_\epsilon)$, $\sigma_\epsilon = 1$, which causes apparent changes in development of all variables, Y_t, π_t, r_t, i_t, but in Y_t, in particular, which is to be expected naturaly. In Fig. 3, we use the same data, again, except the exogeneous supply shock $\eta_t \in N(0, \sigma_\eta = 1)$, applied on π_t, within the Philips curve. In Fig. 4, we show combined effects of both exogeneous shocks $\epsilon_t \in N(0, \sigma_\epsilon = 1), \text{and}\, \eta_t \in N(0, \sigma_\eta = 1)$, applied on y_t, π_t, respectively.

In Fig-s. 5, 6, 7, we use the same data and simple shock on Y_t, $\epsilon_2 = 1$, as in Fig. 1, but with different versions of generalized expectation of inflation, i.e. calculation of $E_t\pi_{t+1}$, using Eq. (10), with $\kappa_b = 0,\ 0.5,\ 0.95$, respectively. Comparing with the results presented in Fig.1, we conlude the evident influence in case of $\kappa_b = 0, \text{or}\, 0.5$, but in the case of $\kappa_b = 0.95$, the production gap $(Y_t - \bar{Y}_t)$ is practically the same as in the Fig. 1, and also, the differences $(\pi_t - E_t\pi_{t+1})$, are almost negligible, while in Fig. 1, is $E_t\pi_{t+1} = \pi_t$, as a consequence of the Eq. (4).



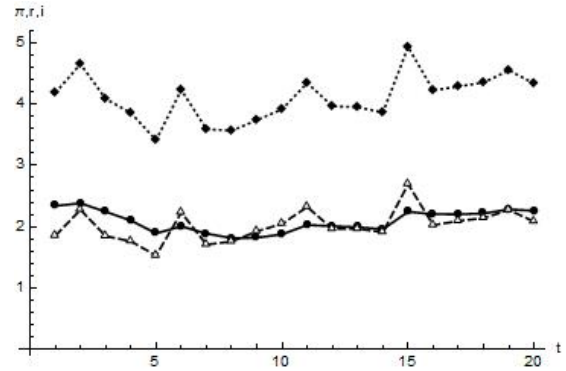
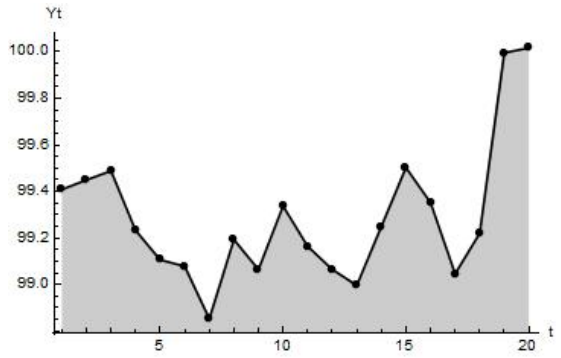


Figure 2: Left panel: Product Y_t, with shock $\epsilon_t \in N(0,1)$. Right panel: π_t (full line), r_t (dashed), i_t (dotted).



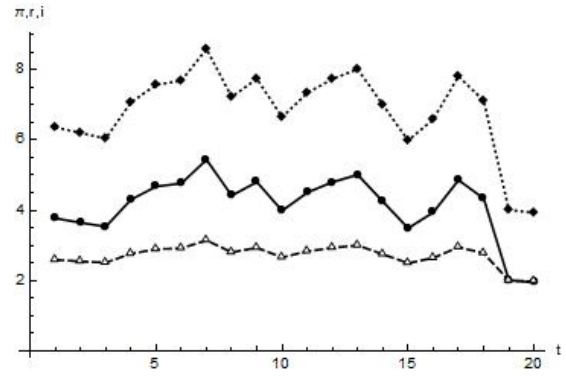
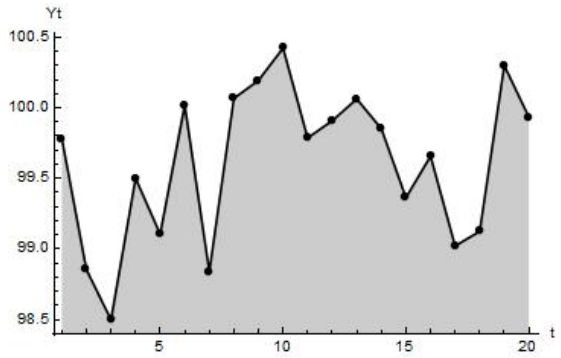


Figure 3: Left panel: Product Y_t. Right panel: π_t (full line), with shock $\eta_t \in N(0,1)$, r_t (dashed), i_t (dotted).



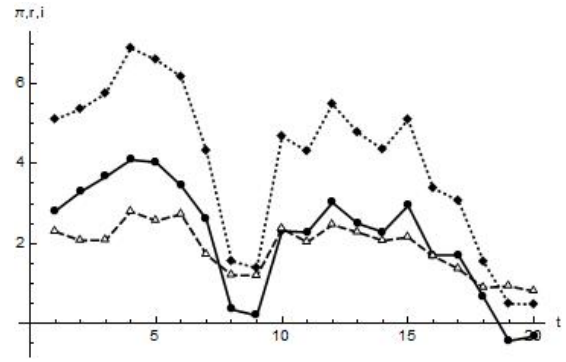
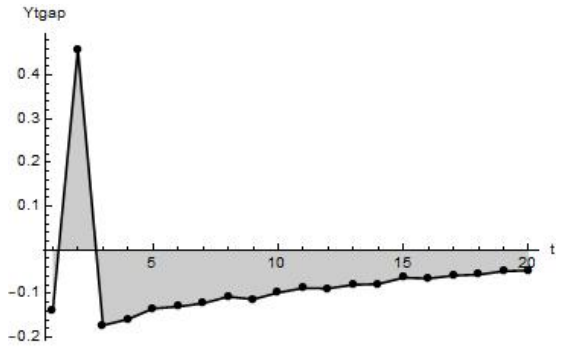


Figure 4: Left: Product Y_t, with shock $\epsilon_t \in N(0,1)$. Right: π_t (full line), with shock $\eta_t \in N(0,1)$, r_t (dashed), i_t (dotted).



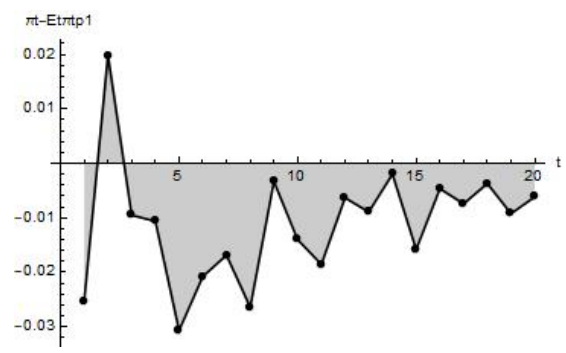


Figure 5: Left panel: Product gap $Y_t - \bar{Y}_t$, with $\epsilon_2 = 1$, and convex expectation $\kappa \in U(0,1), \kappa > 0$.
Right panel: $\pi_t - E_t\pi_{t+1}$.

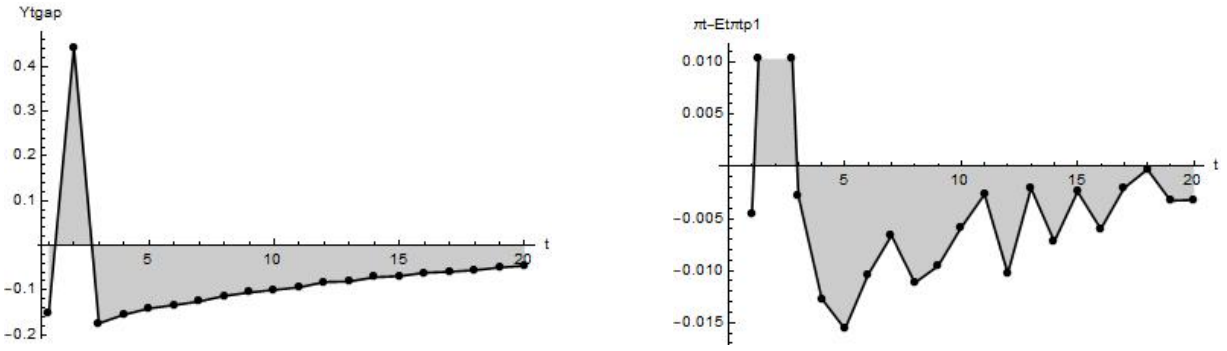


Figure 6: Left panel: Product gap $Y_t - \bar{Y}_t$, with $\epsilon_2 = 1$, and convex expectation $\kappa \in U(0,1), \kappa \geq 0.5$.
Right panel: $\pi_t - E_t\pi_{t+1}$.

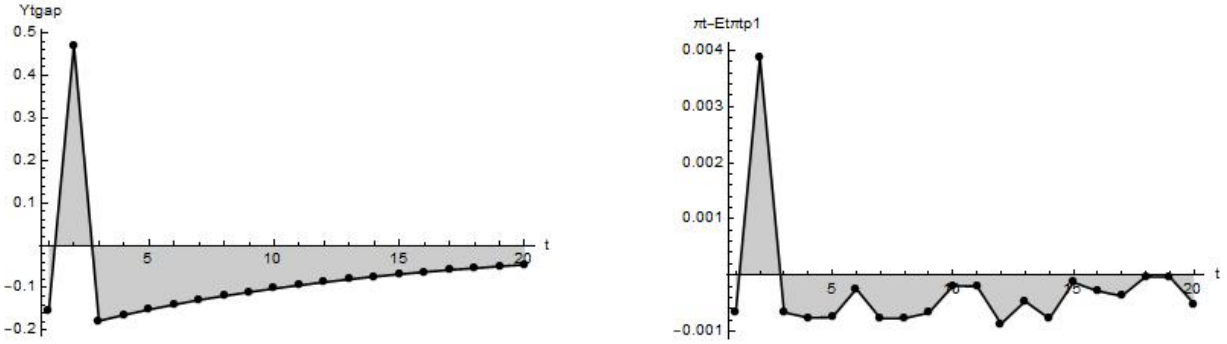


Figure 7: Left panel: Product gap $Y_t - \bar{Y}_t$, with $\epsilon_2 = 1$, and convex expectation $\kappa \in U(0,1), \kappa \geq 0.95$.
Right panel: $\pi_t - E_t\pi_{t+1}$.

4 Conclusions

We have presented formulation of discrete dynamic AD-AS macroeconomic model of Mankiw type discussing all macroeconomic structural equations in detail. Next, we focus our main attention to various general possibilities to handle expected inflation. One of the basic general expectations is formed as convex combinations of past inflations. In particular, we elaborated this type of general expectation of inflation with both deterministic and stochastic setting of convex parameters. We present the numerical implementation of dynamic AD-AS macroeconomic model of Mankiw type both with classical and generalized expectations in Mathematica, and also discuss development of product and inflation under usual type of production shocks.

Acknowledgements

In particular, we acknowledge the financial support of our research by the grant SGS2017-004 Finance and sustainable development from viewpoint of theory and practice, registered in and provided by the Internal Grant System of the University of West Bohemia in Pilsen, Czech Republic.

References

[1] Feuerriegel, S., and Gordon, J. (2019). News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. *Eur. J. of Operational Research*, 272, 162-175.

[2] Flaschel, P., Frank, R. & Semmler W. (1997). *Dynamic Macroeconomics*. Cambridge: MIT Press Ltd., Cambridge, Massachusets, US.

[3] Ljungquist, L. & Sargent, T.J. (2004). *Dynamic Macroeconomic Theory*. 2-nd Ed. Cambridge: MIT Press Ltd., Cambridge, Massachusets, US.

[4] Mankiw, N.G. (2009). *Macroeconomics*. 7-th Ed. New York: Worth Publishers, 41 Madison Avenue, NY 10010, US.

[5] Wieland, V., Cwik, T, Mueller, G. J., Schmidt, S, and Wolters, M. (2012). A new comparative approach to macroeconomic modelling and policy analysis. *J. of Econ. Behavior & Organization*, 83, 523-541.

[6] Wuthisatlan, R., and Thanetsunthorn, N. (2019). Teaching macroeconomics with data: Materials for enhancing students' quantitative skills. *Int. Review of Econ. Education*, 30, 100151.

Inverse data envelopment analysis models: comparison of multi-objective optimization approaches

Josef Jablonský[1]

Abstract. In traditional data envelopment analysis (DEA) models, the efficiency of a decision making unit (DMU) is evaluated relatively with respect to other elements of the set of homogeneous units based on the values of multiple inputs and multiple outputs. These models assign maximum efficiency score to the units lying on the efficient frontier derived by the model. The inefficient units have worse efficiency score that the efficient ones. Inverse DEA models deal with different tasks. Typical problems, in this case, are as follows. If the inputs of a unit under evaluation increase (decrease) how much the outputs have to decrease (increase) to maintain the current level of efficiency, or if the outputs decrease (increase) how much the inputs must change to efficiency score remains unchanged. The paper discusses various approaches to solve these problems and compares the results for different groups of DEA models. They are typically based on multiple objective linear programming methodology. A simple numerical example illustrates the presented approaches and their results.

Keywords: Data envelopment analysis, multi-objective optimization, goal programming, efficiency

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Traditional DEA models were introduced by Charnes et al. [3] and further extended by Banker et al. [2] and by many other researchers. They deal with efficiency analysis of a set of homogeneous DMUs that transform multiple inputs into multiple outputs. In traditional DEA models, the efficiency score θ_q of the DMU_q is defined as the weighted sum of outputs divided by the weighted sum of inputs:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{qk}}{\sum_{j=1}^{m} v_j x_{qj}}, \tag{1}$$

where u_k, $k = 1, \ldots, r$ is the positive weight of the k-th output, v_j, $j = 1, \ldots, m$ is the positive weight of the j-th input, and x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ and y_{ik}, $i = 1, \ldots, n$, $k = 1, \ldots, r$, are non-negative values of the j-th input and the k-th output for the DMU_i. The optimization model for efficiency evaluation of the DMU_q maximizes the efficiency score (1) under the assumption that efficiency scores for all other units of the set are lower or equal than 1:

$$\frac{\sum_{k=1}^{r} u_k y_{ik}}{\sum_{j=1}^{m} v_j x_{ij}} \le 1, \qquad i = 1, \ldots, n, \tag{2}$$

and the weights of the inputs and outputs are strictly positive. This optimization model is not linear in the objective function but can be moved to a linear problem using Charnes-Cooper transformation easily. This linearized problem in its output orientation is often denoted as multiplier CCR model, and its formulation follows:

Minimize

$$\sum_{j=1}^{m} v_j x_{qj}$$

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, Praha 3, Czech Republic, e-mail: jablon@vse.cz

subject to (3)

$$\sum_{k=1}^{r} u_k y_{ik} - \sum_{j=1}^{m} v_j x_{ij} \le 0, \qquad i = 1,...,n,$$

$$\sum_{k=1}^{r} u_j y_{qk} = 1,$$

$$v_j \ge \varepsilon, \qquad j = 1,...,m,$$

$$u_k \ge \varepsilon, \qquad k = 1,...,r.$$

Envelopment form of the CCR output-oriented model is derived from the dual of model (3):

Maximize θ_q

subject to (4)

$$\sum_{i=1}^{n} x_{ij}\lambda_i \le x_{qj}, \qquad j = 1,...,m,$$

$$\sum_{i=1}^{n} y_{ik}\lambda_i \ge \theta_q y_{qk}, \qquad k = 1,...,r,$$

$$\lambda_i \ge 0, \qquad i = 1,\ldots,n.$$

The optimal objective function value of model (4) θ_q^* is greater than 1 for inefficient units and equals 1 for the units weakly or fully efficient.

Traditional DEA analysis consists in solving models (3) or (4) which leads to deriving efficiency scores for all units of the set and other characteristics that contribute to a more in-depth analysis of the problem. In 2000, Wei et al. [10] introduced inverse DEA problems. They deal with answering the following question: How much the outputs have to decrease (increase) to maintain the current level of efficiency, if the inputs of a unit under evaluation increase (decrease). More generally, they deal with the modification of inputs and/or outputs to reach a specified level of efficiency of the unit under evaluation under the assumption that the inputs and outputs of all other DMUs remain unchanged.

The main aim of the paper is to propose new models for dealing with inverse DEA models based on goal programming methodology and compare their results. It is organized as follows. Section 2 contains the basic formulation of inverse DEA models and a brief literature review. Section 3 presents various multi-objective approaches to solving inverse DEA models with special attention to the application of goal programming methodology. The results of the presented approaches are illustrated on a real example with 20 DMUs, two inputs and three outputs. The final section discusses the results and concludes the paper.

2 Inverse DEA models

Many researchers in the past have studied inverse DEA models. After the introductory work Wei et al. [10] in 2000 where the inverse DEA models were formulated for the first time, Jahanshahloo et al. [7] proposed models based on inverse DEA to estimate the best level of inputs and to identify the extra inputs of a DMU when its outputs and efficiency is modified. They provided sufficient and necessary conditions for input estimation under the same efficiency level. Ghobadi and Jahangiri [4] discussed current theoretical results and methodological issues of inverse DEA models and extended the theory by theorem, which provides a sufficient condition for efficiency maintaining in the presence of fuzzy data. Lim [9] deals with inverse DEA problems considering expected changes in the production frontier in the future by integrating the inverse optimization problem with a time series application of DEA. The research in inverse DEA models is not finished. There are still published various studies using this theoretical concept. A good example is the paper of Amin et al. [1] that is focused on target setting in mergers using goal programming and inverse DEA analysis.

Let us suppose that the efficiency score of the unit under evaluation DMU_q is ϕ_q^* with the inputs x_{qj}, $j = 1,\ldots, m$, and outputs y_{qk}, $k = 1,\ldots, r$, and $\bar{x}_{qj} = x_{qj} + \alpha_{qj}$, $j = 1,...,m$, are the modified values of inputs. The question is how to change the outputs to maintain the current level of efficiency or to reach the desired level of efficiency given by the decision maker. Let us denote the new output values as $\bar{y}_{qk} = y_{qk} + \beta_{qk}, k = 1,...,r$, where β_{qk} are values expressing the necessary increase or decrease of outputs to ensure the efficiency level ϕ_q^*. The mathematical formulation of the output-oriented inverse DEA model with the assumption of constant returns to scale is as follows:

$$
\begin{aligned}
&\text{Vector max} && (\beta_{q1}, \ldots, \beta_{qr}) && \\
&\text{subject to} && \sum_{i=1,i\neq q}^{n} x_{ij}\lambda_i + \lambda_q(x_{qj}+\alpha_{qj}) \le x_{qj}+\alpha_{qj}, && j = 1,...,m, \qquad (5)\\
& && \sum_{i=1,i\neq q}^{n} y_{ik}\lambda_i + \lambda_q(y_{qk}+\beta_{qk}) \ge \phi_q^*(y_{qk}+\beta_{qk}), && k = 1,...,r,\\
& && y_{qk}+\beta_{qk} \ge 0, && k = 1,...,r,\\
& && \beta_{qk} - \textit{free}, && k = 1,...,r,\\
& && \lambda_i \ge 0, && i = 1,\ldots,n.
\end{aligned}
$$

Similarly, the input-oriented inverse model is as follows:

$$
\begin{aligned}
&\text{Vector min} && (\alpha_{q1}, \ldots, \alpha_{qm}) && \\
&\text{subject to} && \sum_{i=1,i\neq q}^{n} x_{ij}\lambda_i + \lambda_q(x_{qj}+\alpha_{qj}) \le \theta_q^*(x_{qj}+\alpha_{qj}), && j = 1,...,m, \qquad (6)\\
& && \sum_{i=1,i\neq q}^{n} y_{ik}\lambda_i + \lambda_q(y_{qk}+\beta_{qk}) \ge y_{qk}+\beta_{qk}, && k = 1,...,r,\\
& && x_{qj}+\alpha_{qj} \ge 0, && j = 1,...,m,\\
& && \alpha_{qj} - \textit{free}, && j = 1,...,m,\\
& && \lambda_i \ge 0, && i = 1,\ldots,n,
\end{aligned}
$$

where θ_q^* is the efficiency score of the DMU_q derived using input-oriented CCR model or any other efficiency score specified by the decision maker. Model (6) answers the question how to reduce the inputs when outputs increase to their new level $\bar{y}_{qk} = y_{qk} + \beta_{qk}$. Variables α_{qj} and β_{qk} in models (5) and (6) have to be free because they need not be non-negative in case ϕ_q^* or θ_q^* is specified by the decision maker and not as the optimal solution of traditional DEA models.

Models (5) and (6) are not linear in the terms $\lambda_q(x_{qj}+\alpha_{qj})$ and $\lambda_q(y_{qk}+\beta_{qk})$ but they may be omitted because $\lambda_q = 0$ for the inefficient units and the same holds for efficient units if a super-efficiency DEA model is applied. By omitting the above terms, the models become linear.

3 Multi-objective optimization models

Models (5) or (6) are typical multi-objective linear programming (MOLP) problems that can be solved using traditional approaches like lexicographic optimization, goal programming models, etc. All of them are based on the transformation of original models to models with single objective functions. In our numerical experiments further, we will use universal principles known from goal programming methodology. Goal programming is an optimization approach that minimizes deviations (positive or negative or both) from goals specified by decision makers instead of direct optimization of particular objective functions. The preferences among goals may be expressed in different ways – lexicographically, by weights of the goals, or the combination of both approaches. Even though the methodology of goal programming is not new (see Lee [8]), it belongs to one of the most often used principles in various multi-objective applications. The following approaches for solving problems (5) and (6) can be applied:

- *Lexicographic optimization*. Let us suppose that the outputs in model (5) and inputs in model (6) are ranked according to their importance, i.e. the first output is the most important. Similarly, the first input in model (6) is of the highest importance. Then, the multi-objective function of models (5) and (6) moves to the single objective function:

Maximize	β_{q1}	for model (5), or
Minimize	α_{q1}	for model (6).

If the models with these objective functions have a unique optimal solution, then the process is stopped otherwise continues with the objective functions of the secondary importance with additional constraints $\beta_{q1} = \beta_{q1}^*$ or $\alpha_{q1} = \alpha_{q1}^*$ where β_{q1}^* and α_{q1}^* are optimal objective values obtained from the first lexicographic step.

- *Maximization/minimization of the weighted sum of relative deviations*. We can consider α_{qj} and β_{qk} variables as deviations from the original values of inputs and outputs. Then the objective functions of models (5) and (6) can be replaced by the following ones:

Maximize $\sum_{k=1}^{r} w_k \frac{\beta_{qk}}{y_{qk}}$ for model (5) or

Minimize $\sum_{j=1}^{m} v_j \frac{\alpha_{qj}}{x_{qj}}$ for model (6),

where w_k and v_j are weights expressing the importance of the outputs/inputs. Please note that the deviations are considered relatively to the original values of outputs/inputs.

- *Maximization/minimization of minimum/maximum relative deviations*. A modified complete model (5) that maximizes the minimum relative deviation is formulated as follows:

Maximize D

subject to
$$\sum_{i=1,i\neq q}^{n} x_{ij}\lambda_i + \lambda_q(x_{qj} + \alpha_{qj}) \leq x_{qj} + \alpha_{qj}, \quad j = 1,...,m, \tag{7}$$
$$\sum_{i=1,i\neq q}^{n} y_{ik}\lambda_i + \lambda_q(y_{qk} + \beta_{qk}) \geq \phi_q^*(y_{qk} + \beta_{qk}), \qquad k = 1,...,r,$$
$$y_{qk} + \beta_{qk} \geq 0, \qquad k = 1,...,r,$$
$$\frac{\beta_{qk}}{y_{qk}} \geq D, \qquad k = 1,...,r,$$
$$\beta_{qk} - \text{free}, \qquad k = 1,...,r,$$
$$\lambda_i \geq 0, \qquad i = 1,\ldots,n.$$

The modification of model (6) for minimization of the maximum relative deviation is formulated similarly.

4 Numerical illustration

The models presented in the previous section will be illustrated in an example that deals with efficiency and performance evaluation of 19 Czech public economic faculties. The data set is taken from Jablonský [6], and it is used rather for illustrative purposes. The following inputs and outputs are taken into consideration.

Inputs:
- The number of academic staff,
- Labour costs in thousands of CZK.

Outputs:
- The number of students,
- The number of graduates,
- RIV points expressing the quantity and quality of research outputs of the faculties.

The complete data set is presented in Table 1 including efficiency scores of the faculties obtained using CCR output-oriented DEA model, i.e. the efficient units have efficiency score equal to 1, the inefficient ones greater than 1. The last column of Table 1 contains super-efficiency scores of the faculties identified as efficient using the CCR model. They are obtained using Andersen and Petersen model, and super-efficiency scores are lower than 1 for output-oriented models. We will illustrate the results of all models on one unit that is not efficient (FIS VŠE with an efficiency score of 1.196).

Faculty	Acad. staff	Labour costs	# of students	#of graduates	RIV points	Eff. score	Super-eff score
FSV UK	137	57831	4105	821	14136	1.000	0.563
EkF JČU	69	26842	1764	522	3362	1.263	
FSE UJEP	67	26246	2200	559	321	1.299	
ESF MU	93	49739	4453	738	3060	1.000	0.978
OPF SU	108	44908	4385	882	2581	1.073	
FE ZČU	61	20063	2312	519	492	1.049	
HF TUL	83	32510	2081	600	2037	1.610	
FES UP	78	35977	2639	556	3925	1.184	
FP VUT	81	30280	2758	821	1961	1.181	
EkF VŠB	175	71448	6539	1701	4995	1.155	
FME Zlín	84	28277	3419	970	2927	1.000	0.753
FFU VŠE	82	42899	3176	805	5392	1.026	
FMV VŠE	172	71074	4713	1301	5252	1.490	
FPH VŠE	106	47113	3778	1022	3435	1.208	
FIS VŠE	100	43880	3332	686	4809	1.196	
NH VŠE	65	28621	2572	462	3936	1.000	0.991
FM VŠE	39	16542	1437	321	1201	1.179	
PEF ČZU	186	121546	9462	2822	5276	1.000	0.761
PEF MU	114	49361	3658	958	5636	1.205	

Table 1 Data set – efficiency and super-efficiency scores

Further, we will analyse the results for FIS VŠE that is inefficient according to the CCR model. Let us suppose that the inputs of this unit increase 110 employees (formerly 100) and labour costs are newly increased to 45000 (formerly 43880). Table 2 contains results of the following models for two cases: a) The efficiency score of the unit under evaluation remains unchanged, and b) The faculty increases its efficiency score and becomes efficient. We suppose that in both cases, all three outputs must either increase or remain unchanged.

- Lexicographic models with highest preferences set to the first, second, and third output, respectively.
- The weighted sum of relative deviations (weights are identical for all outputs).
- Maximization of the minimum relative deviation.

Table 2 presents the target values of all three outputs and the sum of relative deviations and maximum relative deviation for all instances. The first part of Table 2 contains results of the case when the efficiency score is on the current level (1.196), the second part presents results how to increase the outputs to reach the efficient frontier.

Model	# of students	#of graduates	RIV points	Sum dev. [%]	Max. dev. [%]
Lex O1	3679.51	686	4809	10.43	10.43
Lex O2	3332	976.07	4809	42.28	42.28
Lex O3	3332	686	6485.63	34.86	34.86
Sum of dev.	3332	835.27	6319.57	53.17	31.41
Maximin	3600.2	763.09	5196.09	27.34	11.24
Lex O1	4573.13	686	4809	37.25	37.25
Lex O2	3332	1255.24	4809	82.98	82.98
Lex O3	3332	686	10610.8	120.64	120.64
Sum of dev.	3476.07	752.35	10204.1	126.18	112.19
Maximin	4305.84	912.66	6214.53	91.49	33.04

Table 2 Target values of outputs and results of optimization models

5 Conclusions

One of the important tasks in the application of DEA models is the determination of target values of inefficient units that ensure the desired efficiency level. DEA models derive these target values in a certain way corresponding to the model. Nevertheless, there are many other possibilities how to change inputs or outputs of the unit under evaluation that ensure maximum efficiency (reach the efficient frontier) or at least a given efficiency level (efficiency score). Inverse DEA models with multi-objective techniques can contribute to finding such target values that will satisfy better the preferences of decision makers. This paper summarizes some of the possible approaches on using multi-objective models for finding target values in DEA models. Future research will be focused on analysis of different sources of non-homogeneity in efficiency and performance analysis and formulation of original models for their solving.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 19-08985S - *Models for efficiency and performance evaluation in non-homogeneous economic environment*.

References

[1] Amin, G.R., Al-Muharrami, S. & Toloo, M. (2019). A combined goal programming and inverse DEA method for target setting in mergers. *Expert Systems with Applications*, 115(1), 412–417.
[2] Banker, R.D., Charnes, A. & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
[3] Charnes, A., Cooper, W.W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
[4] Ghobadi, S. & Jahangiri, S. (2013). Inverse DEA: Review, extension and application. *International Journal of Information Technology and Decision Making*, 14(4), 805–824.
[5] Jablonský, J. (2012). Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal of Operations Research*, 20(3), 435–449.
[6] Jablonský, J. (2013). Analysis of research and teaching efficiency of economic faculties in the Czech Republic. In: *Efficiency and responsibility in education*. Praha, 233–239.
[7] Jahanshahloo, G.R., Hosseinzadeh Lotfi, H., Shoja, N., Tohidi, G. & Razavyan, S. (2004). Input estimation and identification of extra inputs in inverse DEA models. *Applied Mathematics and Computations*, 156(2), 427–437.
[8] Lee, S.M. (1972). *Goal programming for decision analysis*. Auerbach Publ., Philadelphia.
[9] Lim, S.J. (2016). Inverse DEA with frontier changes for new product target setting. *European Journal of Operational Research*, 254(2), 510–516.
[10] Wei, Q., Zhang, J. & Zhang, X. (2000). An inverse DEA model for inputs/outputs estimate. *European Journal of Operational Research*, 121(1), 151–163.

Property and Violent Crime: Evidence from the Czech Republic

Jakub Škrobánek[1], Petra Tomanová[2]

Abstract. This paper is focused on analysis of monthly time series of registered violent crime in four selected Czech regions between years 2009 – 2016. The data were obtained from the Police of the Czech Republic and the Police Presidium. Due to the difficulty of obtaining data from the Police Presidium, the dataset and the subsequent analyses of the Czech criminality can be viewed as unique. The main purpose of this paper is to examine state and trend development of the overall violent crime in the Czech Republic with a main focus on Moravian-Silesian Region, Ústí nad Labem Region, Central Bohemian Region and Prague using methods based on time series decomposition. The average violent crime in the monitored period was roughly 50% higher in the Moravian-Silesian Region and Ústí nad Labem Region compared to Prague and the Central Bohemian Region which can be explained by unemployment. Moreover, we have found that there is a downward trend in crime throughout the year.

Keywords: monthly time series, trend, criminality

JEL Classification: C22
AMS Classification: 62J05

1 Introduction

In this paper, the monthly time series of registered violent crime in four selected Czech regions between years 2009 – 2016 are analyzed. The time series of criminal data exhibit strong seasonality pattern and a noticeable trend which is well documented in the literature, e.g. [4]. In this paper, both seasonality and trend will be thoroughly investigated using Czech data. The regions for this analysis were selected based on more characteristics. One of them is the unemployment rate, which could play a significant role in the case of crime. Hypothesis is that crime in regions with lower unemployment is lower than in regions with higher unemployment. The regions are also very different in their area size. On the contrary, the number of inhabitants in them is very similar, which makes it easy in terms of comparison of absolute indicators.

For individual regions data are available from 2009 to 2016 in the database of the Police of the Czech Republic [6]. Data from previous years are available as well; however, the shorter period is used since the methodology of territorial recording of crime in regions was changed by the Police of the Czech Republic in 2008 and thus, those data are not comparable.

In Section 1.1 and 1.2, the basic indicators in criminology and main problems of crime data processing are discussed. In Section 2 and 3, used statistical methods and data preprocessing are described respectively. Finally, in Section 4, we deal with the analysis itself. Section 5 concludes.

1.1 Indicators used in crime analysis

One of many terms used in relation to crime is the so-called *range* of crime. This is the number of acts that can be judged by criminal law as crime committed in a defined area, usually per year. This indicator is expressed in absolute terms. In this paper, we mainly utilize this indicator, and consider one month's time interval.

If we monitor crime in individual regions, which we then compare with each other, demographic influences such as the number of inhabitants must also be considered. When interpreting absolute numbers, the actual state of crime in a given region may be distorted. Therefore, the term *intensity* is often used. It is the proportion of the number of crimes in a given territory and the population in that territory [5].

1.2 Registered and latent crime problem

In professional criminology publications, but also in the media, we can encounter the above-mentioned terms, such as the *range* or *intensity* based on statistical data that are processed by official crime agencies (e.g. police authorities, courts and prosecutor's offices). If we work with such data (as in this paper), we must bear in mind

[1] University of Economics, Prague, Department of Statistics, W. Churchill 1938/4, 130 67 Prague 3, skrj02@vse.cz.
[2] University of Economics, Prague, Department of Econometrics, W. Churchill 1938/4, 130 67 Prague 3, petra.tomanova@vse.cz.

that these are not numbers describing actual crime, but only that part of the crime that the institutions have already detected. The difference between actual and registered crime is called *latent* or hidden crime. Also, in practice, cases that are not criminal offenses are included in the set of registered crime. For example, the police may misclassify the offense [5].

If we want to assess real crime, we need to know what the ratio between registered and latent crime is. In criminology, it is hypothesized that this ratio is practically unchanged, and therefore the conclusions about criminality can be made based on the available data on registered crime. However, this hypothesis was partly questioned by victimological researches. There are more hypotheses about this ratio, but none of them has yet been fully demonstrated. When interpreting official criminal statistics, it must be assumed that these statistics primarily inform the functioning and effectiveness of crime control institutions. Therefore, a possible rise or fall in crime can be partly explained, for example, by a change in the possibilities and abilities of the police to find out that a crime has been committed. Thus, it may not be a change in actual crime. Only part of the latent crime will move to a registered crime or vice versa [5].

In the light of the above-mentioned facts, it is necessary to complete for the sake of completeness factors influencing the informative value of statistics calculated on basis of data obtained by official institutions dealing with crime. It could be for example demographic influences, legislative changes in criminal law, amnesty decision of the President of the Republic, trends in criminal policy, police focusing on certain places or groups of people, methodological problems and differences in statistical reporting or time lag between committing and registering a crime.

2 Statistical methods

Statistical methods such as trend functions, adaptive methods (moving averages), determination of seasonal deviations, averages and standard deviations are used in this paper for the time series state and trend description. And yet, the chosen methods describe the data more than satisfactorily.

2.1 Seasonality

When analyzing a time series with a period of less than one year, we must often take into account seasonal fluctuations. These fluctuations can be caused, for example, by influences such as climate, increased tourism seasons or weekends. Let´s mark the seasonality length by *s.* In this paper we will work with a length of twelve since we work with time series with monthly periodicity.

Presence of the seasonal component can be tested. For this purpose, we use the seasonality F-test. It is performed on time series cleaned of the trend component, since it is based on a one-way analysis of variance depending on seasonal and random components. This means that it compares seasonal variations with residual variations. In connection with seasonality F-tests, the so-called Bonferroni correction is also used. It is a tool used to perform multiple tests simultaneously. The problem is that as the number of hypotheses tested increases, the likelihood of getting a false positive result is growing. Bonferroni correction is described in detail in [3].

In this paper, Periodogram is utilize as well. As stated in [1], time series analysis using a periodogram means decomposing the time series to sine waves (cycles – periods) with different frequencies. The values of the periodogram of the time series $y_1, y_2, \ldots, y_t$ for *t = 1, 2, ..., N* are defined by the formula

$$I(\omega_j) = \frac{1}{2}(a_j^2 + b_j^2), \tag{5}$$

where

$$a_j = \frac{2}{T}\sum_{t=1}^{N} y_t \sin \omega_j t, \;\; b_j = \frac{2}{T}\sum_{t=1}^{N} y_t \cos \omega_j t \text{ a } \omega_j = \frac{2\pi j}{N}, \; j = 1, 2, \ldots, T/2.$$

ω is given in radians per unit of time, which is the time interval between two adjacent observations.

2.2 Trend modeling techniques

If we use the trend function to describe the trend, we need to base ourselves on several of the following assumptions. The time series is a series of ordered values over time that we obtain by observing a certain indicator at equally long time-intervals. We can write it in form

$$y_t = T_t + a_t, \tag{1}$$

where T_t is a systematic component and represents a deterministic trend that can be expressed as a mathematical function of time variable t, and a_t is a non-systematic component [1]. For the estimation of the parameters of the trend functions in this paper, the least squares method is used.

For description of trend we most commonly use linear and parabolic trend functions in forms

$$T_t = \beta_0 + \beta_1 t, t = 1, 2, \ldots, N, \tag{2}$$

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2, t = 1, 2, \ \ldots, N, \tag{3}$$

where β_0, β_1 and β_2 are unknown parameters.

After we estimate the parameters of the trend model, we should investigate exactly how this model describes the actual course of the time series. Therefore, we examine the difference between the actual values of the time series y_t and the so-called fitted values $\hat{y}_t$, respectively the estimated values of the trend $\hat{T}_t$, expressed by

$$y_t - \hat{y}_t = y_t - \hat{T}_t = \hat{a}_t. \tag{4}$$

We will use these interpolation criteria:

1. *Root mean square error:* for this criterion, the smaller its value, the better the trend model reflects the actual trend;
2. *Adjusted coefficient of determination:* it can be effectively used to compare different trend functions with the different number of parameters for a single time series;
3. *Partial t-tests for model parameters:* the null hypothesis H_0: $\beta_i = 0$ is tested, i.e. the statistical significance of i-th variable;
4. *Residual autocorrelation function:* Coefficients of residual autocorrelation can be understood as linear dependencies of time delayed variables a_t a a_{t-k}. It can be conveniently displayed in correlogram.
5. *Ljung-Box test:* the null hypothesis of no autocorrelation among residuals up to time lag k is tested;
6. *Jarque-Berra test:* the null hypothesis that the residuals are normally distributed is tested. The test statistic compares the skewness and kurtosis of the distribution of the residuals with zero skewness and kurtosis of the normal distribution.

3 Data

The Police of the Czech Republic records violent crime data cumulatively. This means that January data includes all violent crimes registered in January, but for example, March data includes crimes registered from January to the end of March. The data for December is basically data for the whole year.

One of the most suitable way to edit data with the above-mentioned properties is to subtract values of the previous month from values of the current month.

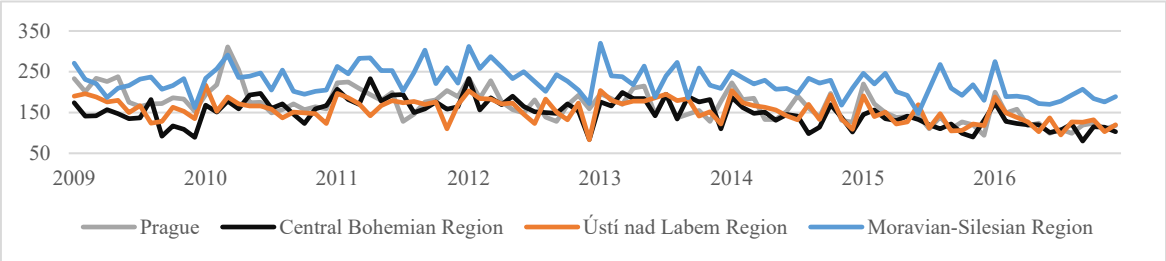


Figure 1 Number of violent crimes in analyzed regions after substraction [6]

Seasonality is slightly visible in the data. Such behavior was expected and is well-documented in the literature, e.g. [2], [4]. Interestingly, the January deviations are often the highest. This could in some way be related to holidays (period after Christmas) or, for example, the amount of alcohol consumed during this period.

4 Crime time series analysis

In this section we will deal with the analysis of time series itself. All tests of statistical hypotheses in this part will be performed at a 5% level of significance unless otherwise stated. Additive decomposition of time series will be considered everywhere as well.

4.1 Descriptive characteristics of violent crime

Table 1 shows the average unemployment rate between 2009 and 2016 calculated as a simple chronological average of the unemployment rate over the reference years, population as of 1 January 2012 and area in square kilometers. Prague, Central Bohemia, Moravian-Silesian and Ústí nad Labem regions are suitable for the comparison since they differ greatly both in terms of unemployment rate and area, which allows us to evaluate crime from several points of view. Conversely, their population numbers are similar (except the Ústí nad Labem Region), which makes the regions comparable in terms of absolute values of recorded crimes.

Region	Average unemployment rate	Population (1.1.2012)	Area (km^2)
Prague	4.15 %	1 241 664	496
Central Bohemian Region	5.92 %	1 279 345	11 016
Ústí nad Labem Region	10.05 %	828 026	5 335
Moravian-Silesian Region	9.07 %	1 230 613	5 428

Table 1 Average unemployment rate, population and area of individual regions between 2009 and 2016 [7], [8]

At first glance, it can be seen in Figure 1 that the number of recorded violent crimes in the Moravian-Silesian Region exceeds the numbers in other regions. It may be related to higher unemployment. The values of the Ústí nad Labem Region cannot be compared with other regions simply by looking at the graph, because it has a significantly different population. However, it can be assured with certainty that the curve of the Ústí nad Labem Region will shift up after certain standardization procedure is performed. The violent crime is recalculated for 100,000 inhabitants.

Region	Average	Average by km^2	Average by 100,000 citizens	Standard deviation	Coefficient of variation
Prague	168	0.34	13.57	39.62	23.5 %
Central Bohemian	150	0.01	11.69	32.77	21.9 %
Moriavian-Silesian	225	0.04	18.29	34.72	15.4 %
Ústí nad Labem	155	0.03	18.75	29.04	18.7 %

Table 2 Descriptive statistics of violent crime in individual regions

From the column of averages by 100,000 inhabitants in Table 2, we can see how the numbers of violence vary considerably for regions with low and high unemployment rates. The average number of recorded violent crimes per square kilometer in Prague is again many times higher than in the remaining regions which is not surprising given the population density.

According to arithmetic averages in Table 2, the highest number of recorded violent crimes is in the Moravian-Silesian Region despite the fact that it has practically the same number of inhabitants as Prague. This observation could be explained by increased police activity in large cities. But there are many other practically undetectable factors that affect violent crime. Also, the area is likely to play a role here. The coefficient of variation is very interesting. It can be generally said that violent crime is higher, however, less variable at the same time in the Moravian-Silesian and Ústí nad Labem regions.

When interpreting variability in time series, it is necessary to take into account that it is influenced by the trend of the time series, which has an effect on the arithmetic mean and thus also on the variance or standard deviation. If we compare two time series with a linear trend and the same random component, with one of the series having a linear trend more severe, the coefficient of variation for both series will vary. Thus, the trend analysis will be performed in Section 4.3 after seasonality adjustment in Section 4.2.

4.2 Seasonality detection

F-test of seasonality using X13ARIMA method is performed for individual regions. The default setting of this method works with multiplicative decomposition. However, due to the visual evaluation of the time series graphs, it will be preferable to use an additive decomposition, as seasonal variations seem to be constant over time. When performing multiple statistical hypothesis tests simultaneously, it is better to use Bonferroni correction. Simultaneous significance levels of 0.05 correspond to individual significance levels of 0.0125. After we performed F-tests, the null hypothesis about the absence of seasonality was rejected in all cases. It can be concluded that a statistically significant seasonal component is present in all monitored time series.

Given that stationary time series adjusted for trend is needed to generate a periodogram, moving averages of length twelve had to be subtracted from the original time series. To identify the period in Table 3, it is sufficient

to look for the highest value calculated according to (5) in the output of the periodogram. In all analyzed time series, the highest periodogram value in the twelve-cycle or six-cycle cycle came out, which can also be seen as the presence of a seasonal component of length twelve. Indeed, the time series may seem to develop as if it had a seasonal component of six. In reality, however, by the coincidence the same seasonal development takes place twice in the same year.

After calculating the absolute seasonal variations, we find that in all regions these deviations are gradually decreasing since January and are the lowest (negative) in December. This is due to an unexpected change in values from December to January.

Prague			Central Bohemian Region		
ω_j	τ_j	$I(\omega_j)$	ω_j	τ_j	$I(\omega_j)$
0.523599	12	25.06731	0.523599	12.00000	14.75295
1.047198	6	7.906281	1.47198	6.000000	3.556767
1.570796	4	6.062342	1.570796	4.000000	6.873776
3.141593	2	9.339286	3.141593	2.000000	9.976190
Moravian-Silesian Region			Ústí nad Labem Region		
ω_j	τ_j	$I(\omega_j)$	ω_j	τ_j	$I(\omega_j)$
0.523599	12	11.44696	0.523599	12	11.5802
1.047198	6	16.32003	1.047198	6	11.16953
1.570796	4	6.749407	1.570796	4	10.3617
3.141593	2	10.73611	3.141593	2	7.122024

Table 3 Periodograms for individual regions

4.3 Trend analysis

After seasonal adjustment, we have chosen models with parabolic trend as the best model in most of the monitored regions on the basis of selected criteria presented in Section 2.1 (see Table 4). In Prague, it appears to be the best model with a quadratic trend, which is same as parabolic, but β_1 is missing. Jarque-Berra tests confirmed for selected parabolic models that residuals are normally distributed.

Figure 2 shows that there has been a roughly three-year growth since 2009 in all regions except Prague. One possible answer seems to be the introduction of the new Penal Code in 2009, which entered into force in early 2010. The new Penal Code, in contrast to the old one, assesses the crime on a formal principle, not a material one. This means that the fullness of all the elements of the actus reus is sufficient to punish the crime. During 2013, there is a relatively sharp increase. It is most visible in the Ústí nad Labem Region. This could be explained by the president's amnesty in January 2013. The re-emergence of the downward trend could then be due to the fact that many recidivists, after their release, again returned to criminal activities and found themselves in prison again.

Region	*Trend*	*$R2_{adj}$*	*t-test*	*RMSE*	*Ljung-Box test*	*ACF*
Moravian-Silesian Region	*Linear*	*0.129*	*0.0002*	*26.41*	*0.0000*	Autocorrelation
	Parabolic	*0.304*	*0.0003* *0.0000*	*23.48*	*0.787*	*OK*
Prague	*Linear*	*0.476*	*0.0000*	*23.03*	*0.391*	Autocorrelation
	Parabolic	*0.495*	*0.7812* *0.0346*	*22.48*	*0.215*	*OK*
	Quadratic	*0.506*	*0.0000*	*22.49*	*0.215*	*OK*
Central Bohemian Region	*Linear*	*0.161*	*0.0000*	*25.86*	*0.0000*	Autocorrelation
	Parabolic	*0.433*	*0.0000* *0.0000*	*21.15*	*0.726*	*OK*
Ústí nad Labem Region	*Linear*	*0.221*	*0.0000*	*19.60*	*0.001*	Autocorrelation
	Parabolic	*0.329*	*0.0147* *0.0001*	*18.09*	*0.052*	Autocorrelation

Table 4 Interpolation criteria

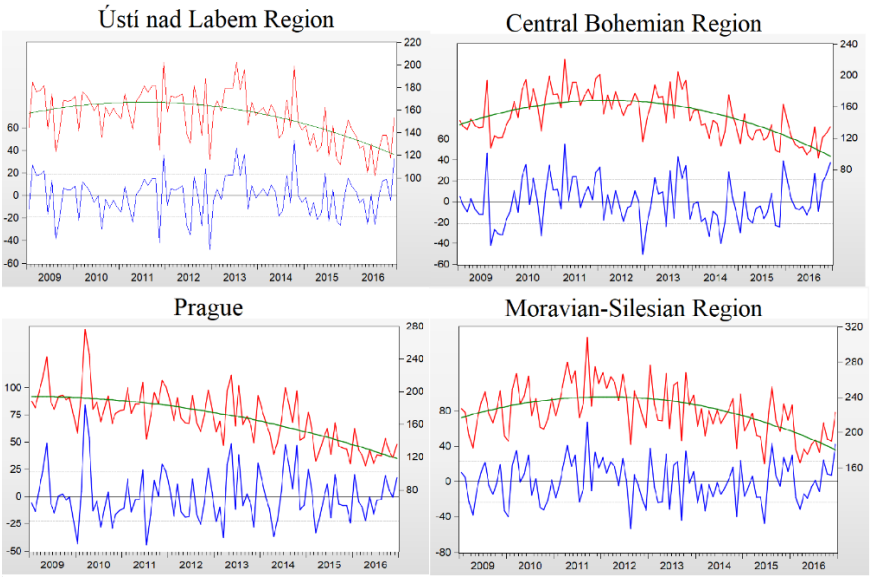


Figure 2 Estimated parabolic trend (green), crime time series (red) and residuals (blue) in all selected regions

5 Conclusion

In this paper, a few important results were found. The average violent crime in the monitored period was roughly 50% higher in the Moravian-Silesian Region and Ústí nad Labem Region compared to Prague and the Central Bohemian Region. This might be caused by a relationship between unemployment and crime since the Moravian-Silesian Region and Ústí nad Labem Region have higher unemployment rate. However, this phenomenon may have been the result of many other influences, which often cannot even be properly captured. Moreover, the results from the analysis document the lowest fluctuations for regions with high unemployment rates. Prague shows the highest variability in the number of recorded violent crimes.

In terms of seasonality, we have found that there is a downward trend throughout the year. In January, the highest positive seasonal variations were detected. By contrast, these deviations were the lowest (negative) in December. All time series exhibit parabolic trend. Trend development seems to be more than optimistic after all. Since 2012, there has been a steady decline in violent crime in all of the selected regions. This can be caused by many socio-economic influences or political and administrative changes.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/53/2019.

References

[1] ARLT, Josef; ARLTOVÁ, Markéta; RUBLÍKOVÁ, Eva, (2002). *Analýza ekonomických časových řad s příklady*. Vyd. 1. Praha: Vysoká škola ekonomická. ISBN 80-245-0307-07.
[2] CHEATWOOD, Derral, (1988). *Is there a season for homicide?*. Criminology. 26.2: 287-306.
[3] CRYER, Jonathan D.; CHAN, Kung-Sik, (2008). *Time Series Analysis With Applications in R*. New York: Springer Science + Business Media. ISBN 978-0-387-75958-6.
[4] LANDAU, Simha F.; FRIDMAN, Daniel, (1993). The seasonality of violent crime: the case of robbery and homicide in Israel. Journal of research in crime and delinquency, 30.2: 163-191.
[5] VÁLKOVÁ, Helena; KUCHTA, Josef, (2012). *Základy kriminologie a trestní politiky.* Vyd. 2. Praha: C. H. Beck. ISBN 978-80-7400-429-2.
[6] Crime time series (seen 02/2018). URL: http://www.policie.cz/statistiky-kriminalita.aspx
[7] Unemployment rates (seen 02/2018). URL: https://www.czso.cz/csu/czso/cr_od_roku_1989_podil_nezamestnanych
[8] Population data (seen 02/2018). URL: https://www.czso.cz/csu/czso/pocet-obyvatel-v-obcich-k-112012-izjb59u5xn

Statistical Inference for Logistic Model: Comparison of Alternative Approaches

Matúš Porázik[1], Petra Tomanová[2]

Abstract. When working with data sets that do not meet ideal model assumptions, we encounter problems with statistical inference. These problems are more prominent in nonlinear regression models where the nonlinear function of the estimated parameters in the form of so-called marginal effects is typically of interest. The aim of the paper is to describe the basic approaches to calculating standard errors and confidence intervals for estimated parameters of logistic regression and then to compare the accuracy of the individual approaches using Monte Carlo experiments on suitably selected examples. The data are generated in a way that when estimating the nonlinear regression models we encounter issues such as the presence of outliers or violation of assumptions imposed on error term.

Keywords: Bootstrap, Logistic model, Logit, Maximum likelihood, Monte Carlo, Quasi maximum likelihood, Sandwich

JEL Classification: C15
AMS Classification: 62F12

1 Introduction

The computational performance of computers is still rising. This allows considerable advances in the field of econometrics. It is no longer impossible to solve complex models in a real time. Computational methods are becoming more and more popular, and more commonly used.

This paper explores methods for solving nonlinear regression models. We can divide these methods for standard error computation into conventional such as Maximum likelihood method, and more robust alternatives such as Quasi-maximum likelihood, Bootstrap and Sandwich.

Conventional methods for modeling of nonlinear models usually rely on strict assumptions. These are often not held in real life situations leading to biased and inconsistent estimates. The most problematic and frequent violations of assumptions are typically the presence of outliers and violation of assumptions imposed on error term. Thus the motivation of this paper is to explore alternative methods that do not rely on such strict assumptions.

We compare four methods, namely Maximum likelihood, Quasi-maximum likelihood, Bootstrap and Sandwich, to estimate Logit and Probit regression models. The data used are generated by the Monte Carlo method which allows us to execute numerous experiments and therefore to obtain a more accurate conclusion.

2 Logit and Probit

2.1 Model specification

The Logit and Probit models work with a binary dependent variable, which indicates whether a modeled event occurs or not. Conventionally the variable is equal to one when the modeled event has occurred and is equal to zero otherwise.

The interest lies in modeling the probability that the event occurs, $y = 1$, conditional on a vector of selected explanatory variables x, i.e. $\mathrm{P}(y = 1|x)$. Therefore we would consider models with binary response for ith observation as follows

$$P(y_i = 1|x_i) = G(y(x_i)),$$

where G is a function acquiring values from zero to one and $y(x_i) = \beta' x_i$, where β is a vector of parameters to be estimated. There are many options for nonlinear functions that can be chosen for G. For the logistic function the choice is

$$G(y(x_i)) = \frac{1}{1 + \exp(-y(x_i))}. \tag{1}$$

[1] University of Economics, Prague, porm00@vse.cz
[2] University of Economics, Prague, petra.tomanova@vse.cz

For probit model: $G(y(x_i)) = \Phi(y(x_i))$, where $\Phi(\cdot)$ denotes the cumulative distribution function $N(0, 1)$ [6].

Latent variable interpretation The logistic regression can be viewed as a method for estimating vector of parameters β that best fit

$$y_i = \begin{cases} 1 & \text{if } \beta' x_i + \varepsilon_i > 0 \\ 0 & \text{if } \beta' x_i + \varepsilon_i \leq 0 \end{cases},$$

where ε is an error distributed by the standard logistic distribution. For Probit model, the standard normal distribution is used instead.

It it worth noticing that the variables y and x are observable. On the contrary, the error term ε is unobservable and consequently, associated latent variable y^* defined for ith observation as $y_i^* = \beta' x_i + \varepsilon_i$ is unobservable as well. Thus, unlike ordinary regression, the vector of parameters β cannot be expressed in terms of observable variables y and x.

Hence, the vector of parameters needs to be estimated based on an iterative search process that finds the maximum of likelihood function which is expressed as a function of values of observable variables y and x. The estimation approach is explained in the next Section 2.2.

2.2 Estimation techniques

In this section we describe four investigated estimation techniques, first being the maximum likelihood method.

Maximum Likelihood The Maximum Likelihood (ML) estimator for the vector of parameters β' is obtained by finding $\hat{\beta}'$ that maximizes

$$L(\beta) = \prod_i^n \left(\frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}\right)^{y_i} \left(1 - \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}\right)^{1-y_i}. \tag{2}$$

In general, the closed-form solution does not exist. Hence the maximum likelihood estimates are obtained by using iterative algorithms such as Newton-Raphson algorithm [5].

Quasi-maximum Likelihood When some of the assumptions for (non-linear) regression are not met, we have to use more robust methods for parameter estimation. The quasi-maximum likelihood method (QML), which operates under weaker assumptions, might be utilized as such a method. QML estimator is the same as ML method (see Equation 2), however, QML does not rely on asymptotic covariance matrix.

Bootstrap As an alternative robust method, Bootstrap can be utilized. Bootstrap is a resampling method with replacement that is, however, computationally intensive. On the other hand, the bootstrap method does not rely on any distributional assumption. For each replication we get a bootstrap sample of size n from the original data set, and consequently, from each jth bootstrap sample we get an estimated value of parameters $\hat{\beta}^{(j)}$. Then the bootstrap standard errors are computed as a standard deviation of $\hat{\beta}^{(j)}$, $j = 1, \ldots, n$. In general, the more bootstrap samples we have, the more precise standard errors we obtain.

Sandwich The other considered alternative robust method is the so-called Sandwich method. The method consistently estimates the covariance matrix of the coefficient estimates in heteroskedastic regression models. The sandwich estimators are of the form

$$(x'x)^{-1} x' \Omega x (x'x)^{-1}.$$

The choice of Ω matrix is somewhat arbitrary, e.g. it might be a function of residuals. In this paper, we utilize the White's estimator.

3 Experimental setup

3.1 Data generating process

For the simulation study, a demonstrative example of a common situation from the banking sector is chosen. We find ourselves in the position of a bank, which needs to decide whether to provide a loan to a customer. Therefore, the dependent variable in our model is binary, and describes whether the customer repays the loan or not.

Let us assume that the bank chooses the following explanatory variables based on the regular variable selection procedure: customer's monthly wages (*wage*), customer's equity (*equity*) and customer's credit flag calculated based on Credit Bureau data (*CBflag*).

Values of the binary dependent variable are simulated as follows using the latent variable reasoning in Section 2. First, ith value of latent variable y_i^* is generated by data generating process

$$y_i^* = 1 + 0.00005 wage_i + 0.00005 equity_i + 0.003 CBflag_i + \varepsilon_i, \tag{3}$$

where ε_i is randomly generated from standard logistic distribution. Then values of the observable binary variable y are calculated as $y_i = I(y_i^* > 0)$, where $I(\cdot)$ is an indicator function.

Customer's monthly wage To obtain values for the customer's monthly wage variable, we randomly generate values from the log-normal distribution which is empirically proved to be a reasonable choice as the distribution of wages. As a mean, we choose the value of 31,646 (precisely $\log(31,646)$ on the log scale) which reflects the average wage in the Czech Republic in 2017 [3]. And as a standard deviation, we choose a reasonable value of $\log(1.5)$ on the log scale.

Customer's equity To obtain values for the customer's equity variable, we randomly generate values from the normal distribution. As a mean, we choose value of 321,000 which reflects the average equity of a citizen from the Czech Republic in 2017 [1]. As a variance, we choose the value 1,000,000. Since we consider indebted customers as well, negative values of the variable are kept.

Customer's credit flag based on Credit Bureau data The customer's credit flag is a binary variable which is equal to 1 for the customer who has repaid all of their past loans in accordance with corresponding contracts and 0 means that he/she failed to meet at least one contract obligation. In the Czech Republic, 11% of customers have not repaid their credit loans [2]. Therefore for each customer, we generate his/her credit flag as follows: credit flag is equal to 0 with probability 0.11 and 1 with probability 0.89.[1]

3.2 Normally distributed error term

Then we slightly change the way how the data is generated. Specifically, the error term of Equation (3) will be generated from the standard normal distribution, $N(0, 1)$. The error term is clearly homoskedastic, however, no longer generated from the logistic distribution. Note, that when ε is standard normally distributed, then the ML estimator using Probit model is unbiased, however this is not true for Logit model.

To show issues with statistical inference under heteroskedasticity, we need to generate the heteroskedastic random error as well. The random variable from Equation (3) is generated from the logistic distribution with variance $1 + exp((equity_i - 321,000)/1,000,000)$. Clearly, the variance is proportional to customer's equity and we encounter the problem of heteroskedasticity.

3.3 Outliers

To investigate the statistical inference under presence of outliers, those outliers need to be generated as well and included in the dataset. After generating a dataset containing thousand different customers, we replace ten customers' monthly wage by (extremely) high values generated from the exponential distribution: exp(0.000001). Moreover, customer's equities are changed to outlying values as well which are generated from the exponential distribution: exp(0.00000001). As a result, ten customers have an extremely high equity and wage, representing super-rich customers.

[1] The Customer's credit flag is generated independently on wages and equities which is not very realistic. This study analyzes an impact on independent variables, however, the correlation among variables might be easily introduced and the analysis might be repeated.

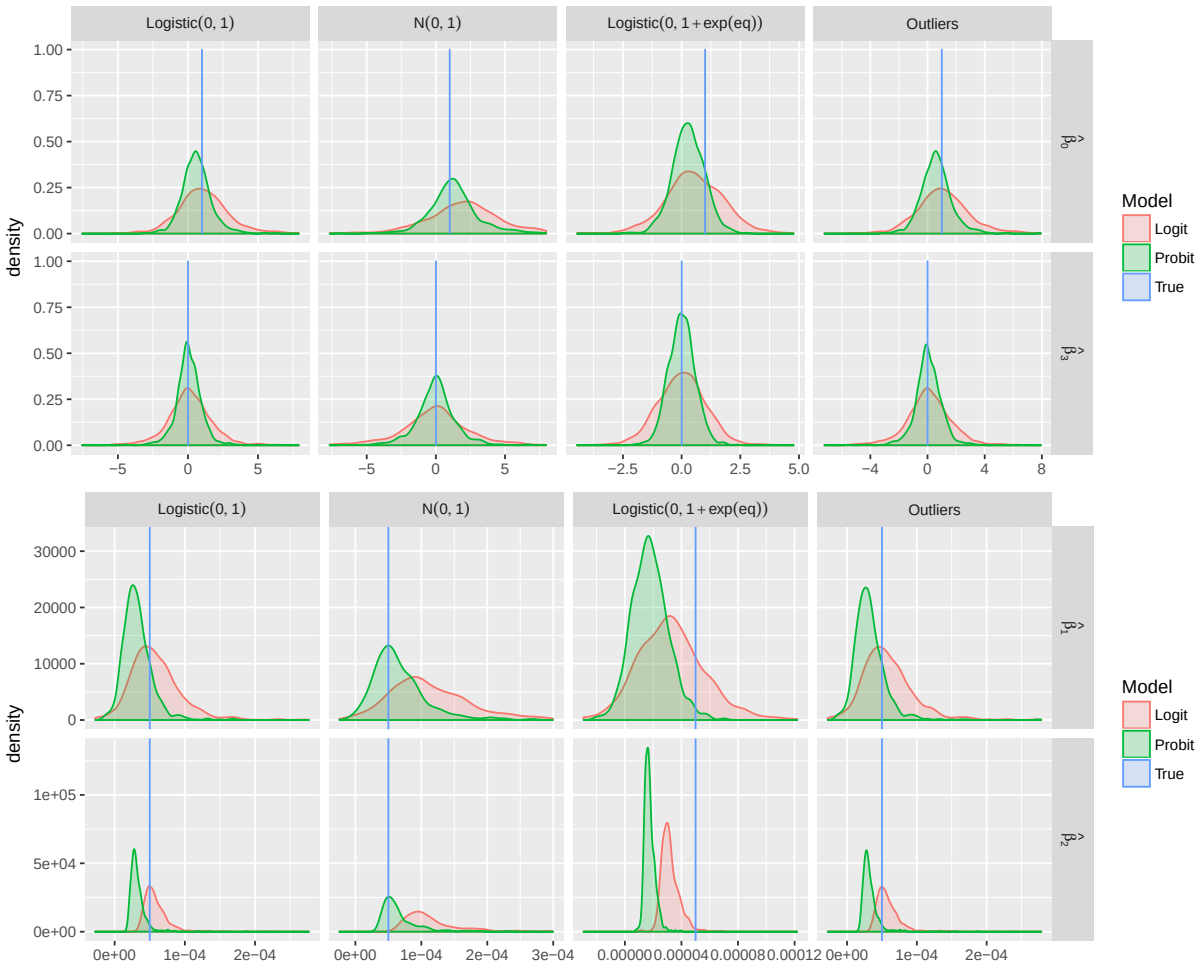


Figure 1 Estimated coefficients of Logit and Probit model based on different scenarios.

4 Simulation study

In this section, four tested scenarios are presented and the expected outcome from the simulation study is stated. Then, results for estimated coefficients and standard errors based on different models and methods are discussed.

4.1 Tested hypotheses

The interest of this paper lies in comparison of estimated parameters of logistic regression models and their standard errors. Since the data are simulated based on specified data generating process, true values of parameters are known.

The following four scenarios are tested:

1. ε_i is generated from standard logistic distribution: $Logistic(0, 1)$;
2. ε_i is generated from standard normal distribution: $N(0, 1)$;
3. ε_i is generated from standard logistic distribution: $Logistic(0, 1 + \exp(eq))$, where eq is standardized *equity* variable;
4. 10 outliers are produced and incorporated into data sample: $wage_i \sim \exp(0.000001)$ and $equity_i \sim \exp(0.00000001)$ for $i = 1, \ldots, 10$.

Under the first scenario, it is expected that the ML estimator gives unbiased estimates for the Logit model and biased estimates for the Probit model (otherwise there will be some unwanted computational issue). Under the second scenario, the situation is reversed: biased estimates for the Logit model and unbiased estimates for the Probit model. Moreover, it is expected that under the first scenario, standard errors and confidence intervals will work well for the Logit model. However, this will no longer be true under the second, third and fourth scenarios.

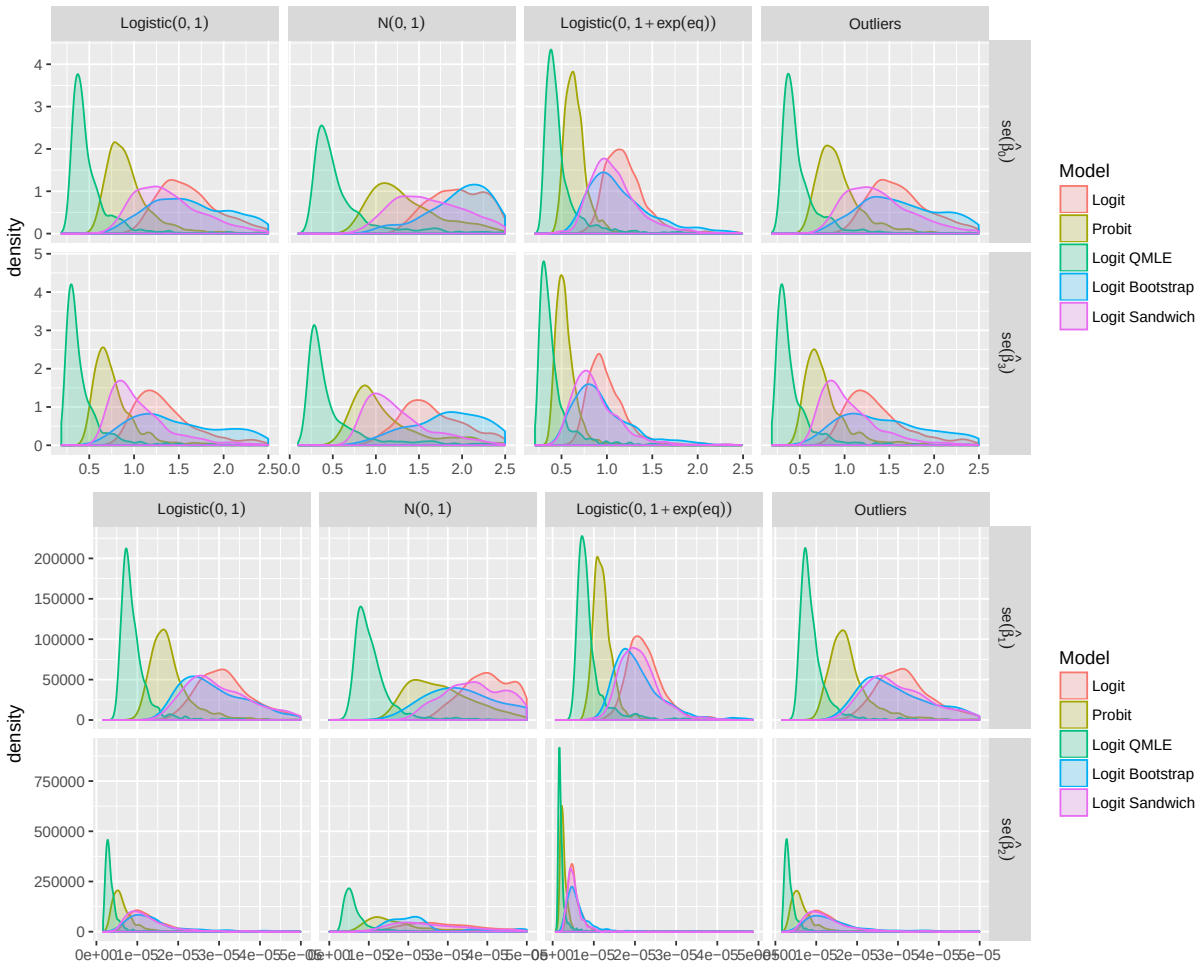


Figure 2 Standard errors of the Logit model based on ML, QML, Bootstrap and Sandwich methods and standard errors of the Probit model based on ML method for different scenarios.

4.2 Results

Each simulation run, 1,000 values for *wage*, *equity* and *CBflag* are generated. Then 1,000 values of ε are generated from corresponding distributions. Moreover, for the fourth scenario, the outliers are added to the sample. Then values of latent variable y^* and consequently values of observable variable y are computed. Finally, coefficients of models are estimated and standards errors and confidence intervals are computed. We execute 1,000 simulation runs.

Estimated coefficients

Figure 1 shows estimated coefficients for both the Logit and the Probit model using the ML estimator. Under the first scenario, all estimates are unbiased for the Logit model as expected. For the Probit model, estimates of β_0, β_1 and β_2 are clearly biased, but β_3 does not seem to be much affected by the wrong choice of the model. Recall that the corresponding variable *CBflag* is a binary variable.

For the second scenario, results are the other way around, i.e. all estimates of the Probit model are unbiased and all estimates except β_3 seem to be biased for the Logit model. Further, Figure 1 shows that all estimates except the one for β_3 are biased due to the heteroskedastic random error (third scenario). The bias is highest for *equity* variable which is not surprising since the dependence of random error on the *equity* variable is the source of the issue. Estimates are not heavily effected by the presence of outliers (fourth scenario).

Standard errors of estimated coefficients

Table 1 shows the percentage of cases for which the true value of the parameter lies inside of the 95% profile confidence interval computed based on ML estimates of Logit model. Profile confidence intervals are more

Parameter	$Logistic(0,1)$	$N(0,1)$	$Logistic(0,1+\exp(eq))$	Outliers
β_0	94.3	90.8	91.3	93.9
β_1	93.3	77.3	85.3	93.2
β_2	93.2	21.0	21.3	92.4
β_3	94.9	93.2	95.0	94.8

Table 1 Percentage of cases when the true value of the parameter lies inside of the 95% profile confidence interval computed based on ML estimates of the Logit model.

complicated to calculate. However, they are definitely preferred over simple model-based intervals [4]. Note that percentages in the first column of Table 1 are not precisely equal to 95 as expected since only 1,000 simulation runs were executed. Three findings are worth to point out: (i) estimates of β_3 (parameter which corresponds to the binary variable) are not heavily affected by different scenarios; (ii) the presence of outliers does not have a big impact on all estimates; (iii) estimates of β_2 (parameter which corresponds to normally distributed variable) are the most affected by different scenarios.

Figure 2 shows computed standard errors of the Logit model based on ML, QML, Bootstrap and Sandwich methods and standard errors of the Probit model based on ML method for different scenarios.

When QML method is used, estimated coefficients remain the same, but standard errors are scaled. Scaling is accomplished by multiplying the square root of the Pearson dispersion by the standard error of the Logit model [4]. Hilbe [4] noted that the resultant standard errors are identical to model-based standard errors if there are no distributional problems with the data. However, for our setup, the scaled standard errors do not seem to work well.

Further, when the error term is normally distributed instead of logistically distributed (second scenario), then the usage of Sandwich standard errors mitigates the impact of the wrong choice of model, i.e. the Logit model is chosen instead of the Probit model.

5 Conclusion

We compared alternative approaches of statistical inference for the Logit and the Probit models using Monte Carlo methods. Our results of the simulation study document an impact of violation of data generating process assumptions of the latent variable on estimated coefficients and standard errors. We showed that the estimates of corresponding binary variable are less sensitive to the wrong model choice. On the other hand, the most affected estimates correspond to the normally distributed variable.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/53/2019.

References

[1] AllianzSE. Global wealth report 2017. https://www.allianz.com/v$_1$506497732000/$media/press/document/AGWR_17-Report_EN.pdf$.

[2] CRIF–CzechCreditBureau. Ohrožené úvěry: české domácnosti v prvním čtvrtletí 2015 splácely své půjčky lépe než před rokem. http://www.cbcb.cz/sqlcache/tz-brki-nrki-20150608.pdf .

[3] CZSO.cz. Počet zaměstnanců a průměrné hrubé měsíční mzdy. https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=vystup-objektpvokc=katalog=30852pvo=MZDD01-Gz=Tw= .

[4] J. M. Hilbe. *Practical guide to logistic regression*. Chapman and Hall/CRC, 2016.

[5] C. Trivedi. *Microeconometrics Methods and Applications*. Cambridge University Press, 2005.

[6] J. M. Wooldridge. *Introductory Econometrics A Modern Approach*. 4th Edition. South-Western, a part of Cengage Learning, Mason, Ohio, 2009.

Tariff analysis in a motor hull insurance portfolio

Adéla Špačková [1],

Abstract. The Non-life insurance pricing is determinated by multiplication of claim frequency and claim severity. The subject of the contribution is tariff analysis, where each empirical model is compared and the categorical model is going to be compare with continous model. The results of this contribution can be interesting and it can contribute to a deeper understanding of this problem and importance of the tariff analysis. All empirical models are estimated on the real-world sample data of czech insurance company collected during the years 2005-2010. Estimation is performed by using generalized linear models. Regression analysis allows the identification of the risk factors and the prediction of the expected frequency of claims given the characteristics of policyholders. It depends on many individual rating factors (e.g. based on individual characteristics of vehicle and driver). The aim of this paper is to find out ideally suited model for estimation premium and point out the importance of tariff analysis.

Keywords: tariff analysis, generalized linear models, claim frequency, individual rating factors.

JEL Classification: C13, G22
AMS Classification: 97M30

1 Introduction

Tariff groups are homogenous groups of insurance contracts, where the insured risk is approximately identical, so in each tariff group is possible to require united insurance rate. Each tariff group corresponds to a certain risk level. At first it is necessary to find out the factors influence the premium and calculate the premium according to the values of these factors. On the basic of these facts insurance company calculate the tariff which has to fulfill the following general principles. The tariff has to be correct as possible and the structure of tariff groups has to be simple (linear function or multiplicative). The aim of this paper is to find out ideally suited model for estimation claim frequency and point out the importance of tariff analysis. The standard method of estimation is Generalized linear models (GLM).

2 Generalized linear models

Generalized linear models (GLM) is a standard estimation method using in insurance practice. Generally, GLM includes three main assumptions:

- A probability distribution have to be form an exponential family;
- A linear predictor is a transform by a link function;
- Link function can be diverse (identity, logarithm, power..).

For the purposes of this paper the logarithm link function is choosen.

$$\ln\left(\frac{\mu}{n}\right) = x'\beta = ln\mu = \ln n + x'\beta \tag{1}$$

where $\ln n$ is called an offset, μ is mean, β is estimated parameter by maximum likelihood method(Cipra 2015).

All probability distribution can be described by the general form as following:

$$f(y) = c(y,\phi)exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\} \tag{2}$$

where θ is the canonical parameter and ϕ is called the dispersion parameter, $a(\theta)$ and $c(y,\phi)$ are determining the actual probability function Gray,Pitts (2012)..

[1] VSB – Technical University of Ostrava, Department of Finance, Sokolská třída 33, 732 00 Ostrava, adela.spackova@vsb.cz.

Insurance premium is given by multiplication of estimated mean of claim frequency and claim severity. Claim frequency is estimated by Negative–binomial distribution and Claim severity by Gamma distribution, as it is shown at Table 1.

Distribution	$\boldsymbol{\theta}$	$\boldsymbol{a(\theta)}$	$\boldsymbol{\phi}$	$\boldsymbol{E(y)}$	$\boldsymbol{V(\mu)=\frac{Var(y)}{\phi}}$
Negative-binomial (μ,κ)	$ln\frac{\kappa\mu}{1+\kappa\mu}$	$-\frac{1}{\kappa}\ln(1-\kappa e^{\theta})$	1	μ	$\mu(1+\kappa\mu)$
Gamma (μ,v)	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{v}$	μ	μ^2

Table 1 Distributions and parameters

2.1 Claim frequency

The random dependent variable is discrete and conditioned by a vector of explanatory variables. (characteristics of risk based on individual characteristics of shareholders) is negative-binomial distributed. The probability of random variable *y* has to be fit into the exponential family framework as following(Long, Freese 2014):

$$\ln\{f(y)\} = y\, ln\frac{\mu}{1+\kappa\mu} - \frac{1}{\kappa}\ln(1+\kappa\mu) = \frac{y\theta - a(\theta)}{\phi} \tag{3}$$

According to Table 1, the dispersion parameter is equal to one and canonical parameter is $\ln\frac{\kappa\mu}{1+\kappa\mu}$.

Mean and variance function is denoted:

$$E(y) = \dot{a}(\theta) = \frac{e\theta}{1-\kappa e^{\theta}} = \mu \tag{3}$$

$$Var(y) = \phi\ddot{a}(\theta) = \frac{e^{\theta}}{(1-\kappa e^{\theta})^2} = \mu(1+\kappa\mu) \tag{4}$$

Where $\dot{a}(\theta)$ and $\phi\ddot{a}(\theta)$ are first and second derivates of $\phi a(\theta)$ with respect to θ.

2.2 Claim severity

In this case the dependent variable is continuous and also conditioned by vector of explanatory variables given by individual characteristics of shareholders. A several literature explains that the most suitable econometrics model permitting classic econometrics of claim severity is the Gamma model The density function is following:

$$ln\{f(y)\} = \left\{\frac{y\theta - a\theta}{\phi}\right\} + (v-1\ln y - ln\Gamma(v) + v\ln v \tag{5}$$

The mean and variance function is denoted:

$$E(y) = a(\theta) = -\frac{1}{\theta} = \mu \tag{6}$$

$$Var(y) = \phi a(\theta) = v^{-1}\frac{1}{\theta^2} = \frac{\mu^2}{v} \tag{7}$$

2.3 Insurance premium

In the case of non-life insurance, the pure premium is given by expected cost of all claims caused by policyholders during the insured period. Mostly, calculation of pure premium includes estimation of the frequency and severity model. Basicly, the mathematical formula can be obtained by multiplication of two estimated components, claim frequency and claim severity:

$$E\left[\sum_{i=1}^{N} sev_i\right] = E(freq_i)\cdot E(sev_i) \tag{8}$$

Standard method of parameters estimation

All models are going to be estimated with standard method called maximum likelihood method. If the maximum likelihood estimation is exponential family distribution, then the probability function is following (Long, Freese 2014):

$$\ell(\beta, \phi) = \sum_{i=1}^{n} \ln f(y_i ; \beta, \phi) = \sum_{i=1}^{n} \left\{ \ln c\ (y_i, \phi) + \frac{y_i \theta_i - a(\theta_i)}{\phi} \right\} \tag{9}$$

Maximization of likelihood called log-likelihood is a logarithm of the likelihood with respect to β_j:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} \tag{10}$$

2.4 Quality testing of selected models

The next step is selection of a suitable model. An important measure of the quality of the model is its predictive ability, therefore the ability to best estimate the value of new or unknown observations. The task of statistics is to choose what might best model, which will use the information contained in the data, but not to overstate the importance of random fluctuations. A simple tool for the selection of a quality model is a statistic known as the deviance.

Deviance

Deviance allows comparison of selected regression model with a saturated model through the logarithm of the likelihood function. In the case of saturated model the estimated values are equal to observed and the value of the logarithm is the highest as possible. Deviance is defined as(Jong, Heller 2008):

$$\Delta = 2(\hat{\imath}^{max} - \hat{\imath}) \tag{11}$$

where $\hat{\imath}^{max}$ is logarithm of saturated model and $\hat{\imath}$ is logarithm of selected model. If the difference between the values of both credibility sufficiently small, then the model can be considered of good quality. Another measure of the quality of the model can be Akaike information criterion.

Akaike information criterion

Information criteria are the measurements that compare the differences between models and serves for the assessment of their relatively quality. Akaike information criterion is defined:

$$AIC = -2lnL + 2k \tag{12}$$

where L is logarithm of the likelihood function k is number of estimated parameters. It should be selected th model, where is the value of AIC is the lowest (Hardin, Hilbe 202).

3 Tariff analysis

According to the framework of the tariff analysis, its going to be estimated a total 3 models. At first it will be estimated the model with continuous data, then continuous variables to be converted to categorical variables. We have to calculate the premium as a function of tariff variable choosen according to Cipra (2006) and Valecký (2015), in accordance with the above mentioned factors, i.e:

$$P = P(x_i \dots, x_k) \tag{13}$$

Our aim is is to find out ideally suited model for estimation premium and point out the importance of tariff analysis. We first have to search for the tariff factors and then construct the tariff. In searching for the factors influencing the tariff we make no assumptions concerning the structure of the tariff.

Every risk can thus be indicated individually using the values of the tariff factors $x_i \dots, x_k$. Thus we get $X = (x_i \dots, x_k)$ as a combination of these values. now you need to clarify the difference between risk and collective premium.

- **The risk premium** is the premium corresponding to the value combination of the tariff factors, and thus it is defined for each value combination separately.

- **The collective premium** is the combination of premium calculated on the different value combination of the different tariff factors. In practice it is usually difficult to use all the k tariff factors. Correspondingly if some of the tariff factors can have many different values, it is in practice preferable to classify the values into a few classes. All the risks belonging to this class have the same premium, which is a collective one. (Ohlsson, Johnanson 2010).

3.1 Data

For the purposes of this paper it was used a random selection of a motor hull insurance portfolio collected during the years 2005-2010 in Czech republic territory. The file contains 18 111 contracts. All regression factors are shown in Table 2.

Variable	Description	Value
Frequency	Claim frequency	0,1,2,3
Severity	Claim severity	3 360- 527 200
Fuel	Type of fuel	1,2,3,4,5
Gender	Gender of driver	0,1
Agecar	Age of car	0-28
Ageman	Age of driver	19-88
Price	Vehicle price	15 000 – 6 704 000
Volumkw	Engine power	2-265

Table 2 Parameters description

In following picture it is shown empirical and predicted frequency by above mentioned variables.

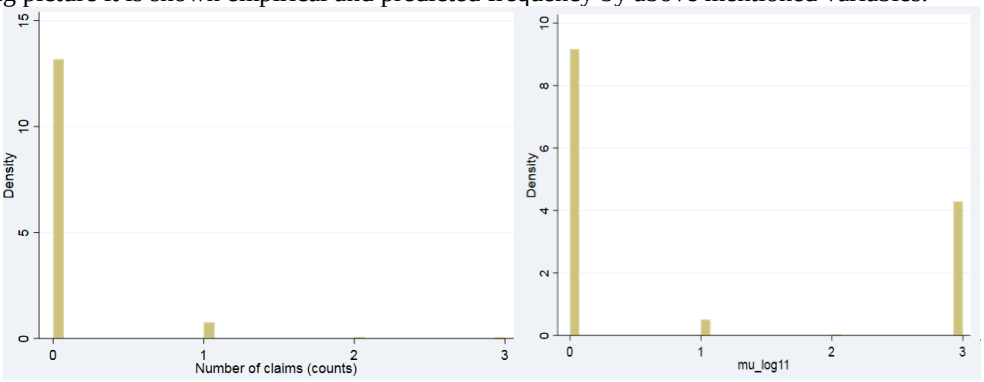


Figure 1 Empirical vs predicted frequency

According to the histogram, it can be argued that the estimated frequency is located between zero and number two, which is comparable with the empirical histogram. In the following picture it is shown deviance of above mentioned model 1.

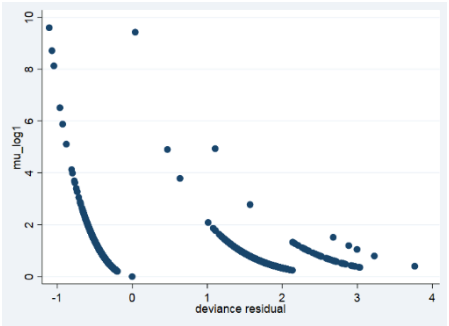


Figure 2 predicted deviance

The value of AIC for model 1 is 5445,135. In the following Table 3 is shown empirical vs. predicted frequency by countinous frequency model1.

Number of claims	Observed	Predicted by Model 1
0	17 045	11,864
1	984	656
2	76	35
3	6	5,556

Table 3 Observed vs. predicted frequency

As we can see, predicted 3 number of claims can be overestimated. In the following Table 4 is shown category of 2 following frequency model. The categories of Model 2 are calculated according subjective opinion, the tariff of model 3 is calculated by quartiles.

Variable	Tariff model 2	Tariff model 3
Frequency	0,1,2,3	0,1,2,3
Severity	3 360- 527 200	3 360- 527 200
Fuel	1,2,3,4,5	1,2,3,4,5
Gender	0,1	0,1
Catgecar	(0-1;2-4;85-9;10-28)	(0-7;8-14;15-21;22-28)
Catgeman	(19-25;26-35;36-50;51-88)	(19-36;37-53;54-70;71-88)
Catrice	(15 000-100 000;-100 001-300 000;300 001-1000 000;1 000 001-6 704 000)	(15 000 –1672 250; 1 672 251-3 344 500; 3 344) (501-5 016 750; 5 016 750 - 6 04 000)
Catvolumkw	(2-10;11-100;101-200;201-265)	29-99; 100- 198; 199-298; 299-426

Table 4 Tariff analysis – frequency

In the following picture we can see the histograms of two predicted frequencies. Model 2 is on the left. And the deviances of both models.

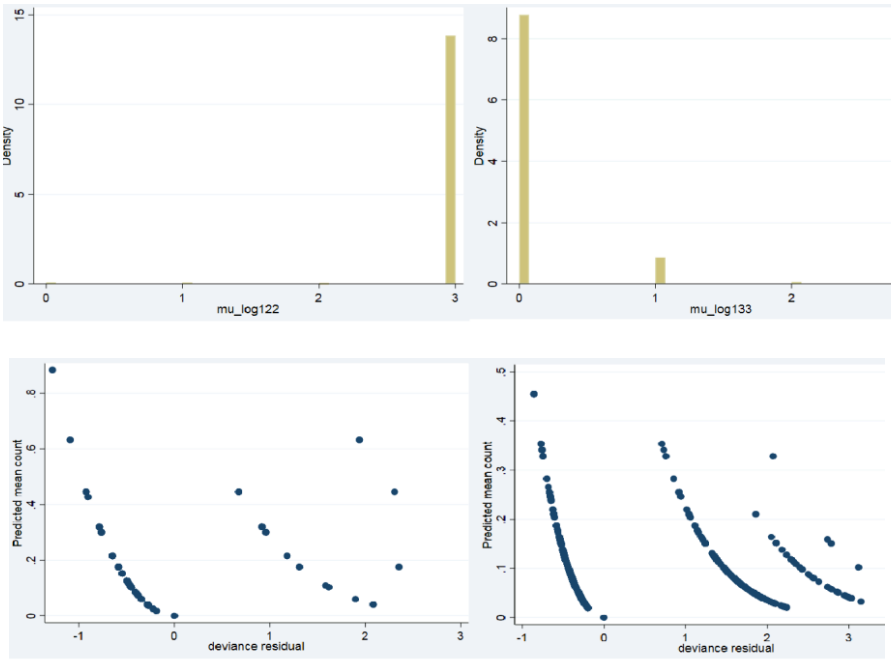


Figure 3 Predicted frequency

As we can see, the deviance of model 3 is lower and also predicted frequency looks like more realistic. The tariff according to quantiles is more appropriate, as we can see the histograms of two predicted frequencies. Model 2 is on the left. and the deviances of both models. In the following table is shown predicted frequency of both models.

Number of claims	Predicted by model 2	Predicted by model 3
0	95	11,350
1	84	1,132
2	12	72
3	17,920	5,557

Table 5 Predicted frequency

The value of Akaike criterion is in the case of model 2 = 5 398,434 and in the case of model 3 it is equal to 216,0583. It can be argued, that model 3 is more appropriate. Now it is time for estimation claim severity. Lets start with continuous model and because model 3 looks like more appropriate, model 2will not be considered

anymore. Observed vs predicted severity b model 1 and 3 is shown in the Figure 4.

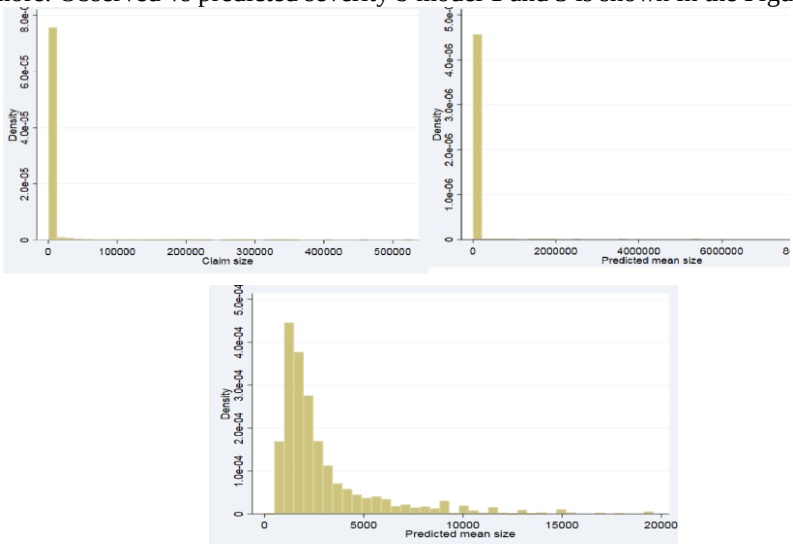


Figure 4 predicted severity

According to the result of AIC the model 3 is more appropriate. In the case of continuous model 1, AIC is equal to 17,504 and in the case of categorical model 3 is 17,447. In the following figure 4 is shown predicted deviance of both models, where the model 1 is on the left.

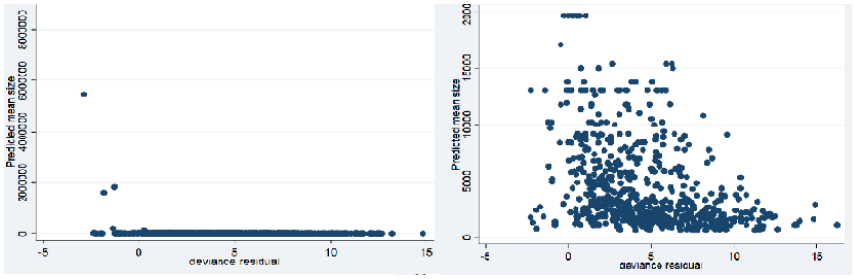


Figure 5 predicted deviance

Categorization of data eased the remote observation, as is evident in Figure 5. In the following the premium. going to be estimated first, for the continuous model, then, for the categorical model. In the end the results will be compared. In the following figure is comparison between empirical and predicted premium by both models.

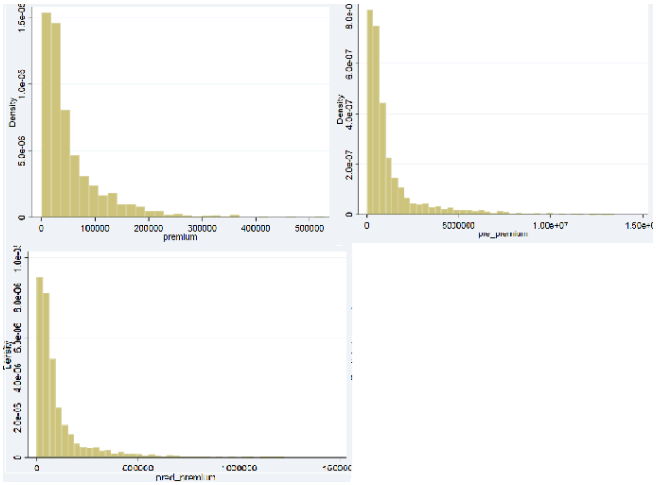


Figure 6 predicted premium

4 Conclusion

Based on the results, it was found that the creation of tariff groups is in the insurance practice very important. The formation of tariff groups can help to calculate cost of risk in individual groups. It has been shown that the

tariffing according to the quartiles is more appropriate. For the reasons that occurred to mitigate the residual component and the absolute value of AIC is the lowest. This paper can contribute to an understanding of a tariff analysis and also as a subject for future research. Tariff analysis as such is extremely interesting and the biggest problem still remains how to properly form a tariff of the group, since their formation significantly affects the calculation of the premium, which is then in each groups slightly different.

Acknowledgements

Supported by the SGS Project VŠB-TU Ostrava SP2019/132 „Finanční rozhodování podniků a finančních institucí za rizika II".

References

[1] CIPRA, Tomáš. *Finanční a pojistné vzorce*. Praha: Grada Publishing, 2006. ISBN 80-247-1633-X.
[2] CIPRA, Tomáš. *Pojistná matematika: teorie a praxe*. 2. aktualiz. vyd. Praha: Ekopress, c2006. ISBN 80-86929-11-6.
[3] CIPRA, Tomáš. *Riziko ve financích a pojišťovnictví: Basel III a Solvency II*. Vydání I. Praha: Ekopress, 2015. ISBN 978-80-87865-24-8.
[4] GRAY, Roger J. a Susan M. PITTS. *Risk modelling in general insurance: from principles to practice*. Cambridge: Cambridge University Press, 2012. ISBN 978-0-521-86394-0.
[5] HARDIN, James W. a Joseph HILBE. *Generalized linear models and extensions*. 3rd ed. College Station: Stata Press, 2012. ISBN 978-1-59718-105-1.
[6] HILBE, Joseph. *Negative binomial regression*. 2nd ed. Cambridge: Cambridge University Press, 2011. ISBN 978-0-521-19815-8.
[7] JONG, Piet de a Gillian Z. HELLER. *Generalized linear models for insurance data*. Cambridge: Cambridge University Press, 2008. ISBN 978-0-521-87914-9.
[8] LONG, J. Scott a Jeremy FREESE. *Regression models for categorical dependent variables using Stata*. Third edition. College Station: Stata Press Publication, 2014. ISBN 978-1-59718-111-2
[9] OHLSSON, Esbjörn a Björn JOHANSSON. *Non-life insurance pricing with generalized linear models*. Berlin: Springer, c2010. ISBN 978-3-642-10790-0.
[10] VALECKÝ, Jiří. Modelling claim frequency in vehicle insurance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis.* 2015. 10 s. ISSN 1211-8516.

Transportation Problem with Time Windows

Dušan Teichmann[1], Michal Dorda[2], Denisa Mocková[3], Alexandra Dvořáčková[4]

Abstract. In the transportation problem we decide about amounts of the commodity transported between the individual sources and destinations so that the total transportation costs are as minimal as possible. However, in real distribution systems which can be optimized by using the transportation problem some time constraints may arise. The time constraints may be defined for the sources or for the destinations; the constraints can be represented for example by opening hours of the sources or the destinations. Another time factor which may play an important role in the optimization process is represented by time durations of different operations which vehicles transporting the commodity must undergo in the distribution system – we can mention for example loading times, unloading times and so on. The basic transportation problem cannot deal with the additional time constraints. In the modified model we do not want to minimize only the total transportation costs but also additional costs resulting from delay times of the vehicles waiting for unloading. Such modification of the transportation problem can be named as the transportation problem with time windows.

Keywords: transportation problem, time windows, binary programming.

JEL Classification: C61
AMS Classification: 90C90

1 Introduction – Problem Solution Motivation

Commodity (product) distribution represents an inseparable part of the supplier chains. Without it, it would not be possible to fulfil the requirements of the end customers who use the given commodity or product for their needs. This is because it is not possible for every commodity (product) consumption location to be located by its mining (production) location. Since commodity or product distribution forms a stage of its cycle, during which its usable characteristics do not change, it is important to make sure it takes place as efficiently as possible. It means that it should take ideally place in an optimal or at least suboptimal manner based on the defined optimization criterion.

Commodity or product distribution manners can differ. Most often, commodity or product distribution is implemented in the form of shuttles or round trips. Distribution in the form of shuttles is mostly used in relations with larger flows of the transported commodities or products. This type of transportation is characteristic for railroad, road, air and maritime transport. Distribution in the form of round trips is used in all other cases. It is mostly used for road transports (even though its implementation for other transportation types, particularly maritime transportation, cannot be ruled out). The selection of the transportation type mostly depends on infrastructure conditions for the distribution using the selected transportation type and, in the case of multiple distribution directions, also on the corresponding vehicle utilization requirements. Round trips are organized when vehicle utilization during shuttles is low.

Our paper addresses optimization of shuttles among multiple sources and multiple customers. The basic optimization instrument for shuttles between sources and customers of a certain commodity or product is the transportation problem. The transportation problem in its basic, the so-called classic form, is mostly used for single-commodity distribution problems, in which customers can be supplied with a given commodity or product from multiple sources (when it is not possible to supply the customers from multiple sources simultaneously, the problem in question becomes an allocation problem). Its use can be also considered even in the case of multicommodity distribution problems, in which simultaneous transportation of shipments of various commodities in the same vehicle is permitted. If shipments of various commodities must be transported separately, we assess the corresponding separate transports as separate distribution problems and we solve them separately.

[1] VSB – TUO, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu, 15/2172, Ostrava-Poruba, dusan.teichmann@vsb.cz.
[2] VSB – TUO, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu, 15/2172, Ostrava-Poruba, michal.dorda@vsb.cz.
[3] CTU in Prague FTS, Department of Logistics and Management of Transport, Konviktská, 20, Praha 1, mockova@fd.cvut.cz.
[4] CTU in Prague FTS, Department of Logistics and Management of Transport, Konviktská, 20, Praha 1, dvoraale@fd.cvut.cz.

The transportation problem in its classic form has it advantages from the perspective of solvability, however, it also has several disadvantages from the perspective of a practical application. It is because, when searching for optimal solutions, it does not consider other types of factors typical for distribution problems in real systems, apart from the source and consumer quantity limitations (source capacities, customer requirements). That is why it can easily happen that the transportation problem in its classic form can offer an optimal solution that will not be acceptable from the perspective of other logistics practice factors.

One of such real limitations not considered in the classic transportation problem is represented by time limitations. Time limitations form a very important aspect, which individual supplier chains must observe. It is because applied logistics often deals with sources that do not operate continuously. Analogically, there are also customers, to whom certain commodities or products cannot be delivered anytime. That is why it makes sense to also consider solutions of transportation problems, in which loading can be conducted at sources and unloading at customer facilities only during time intervals defined in advance.

2 Analysis of the current state of knowledge

A model of the transportation problem in its classic for can be found in a majority of publications devoted to linear programing. Its mathematical formulation was defined in [1]. A very effective method for its solution was proposed in [2].

Apart from the classic variant, other variants of the problem, in which additional limitations are considered, can also be found.

In [3], there is a fixed charge transportation problem, which uses the first approximative approach for its approximate solution, the so-called Balinski Relaxation. The fixed charge transportation problem was also addressed by [4], who solve it utilizing genetic algorithms or the Tabu Search metaheuristic method [5]. One of the latest approaches was published by [6], solution of whose proposes an approach based on a combination of the Lagrangian decomposition and the column generation method.

When we look at [7], we can find another type of the transportation problem – transportation problems with sanctions in the case of unused source capacities and incompliance with customer requirements. Publications [8] and [9] include other types of the transportation problems, such as a transportation problem with variable cost, transportation problem with variable source capacities and customer requirements, transportation problem with limited route capacities, transportation problem with multiple commodities, multicriterial transportation problem and storage transportation problem, in which losses occur on the stored commodities due to storage. Publication [9] also contains the so-called time transportation problem. The variant of the problem minimizes the overall transportation time. It is primarily designated for the transport of perishable goods. The time factor thus plays a completely different role here than in the transportation problem variant that we present in this paper.

Yet another type of the transportation problems is formed by the so-called transportation problems with interim storage facilities. Some examples of such transportation problems are presented in publications [10] or [11]. While publication [10] solves the problem using a linear model, publication [10] presents a solution using an algorithm that looks for the cheapest maximal flow.

Study [12] includes models of the transportation problems, in which the number of the transported units can fall within an interval specified in advance, while the paper [13] contains mathematical models that combine a transportation problem with containers and planning of the circulations of trucks (scheduling vehicles) parked in one depot.

When the transportation expenses develop progressively, we can solve them using the models published in [14]. When the unit transportation expenses develop by degressive steps (degressively stepped costs), the transportation problems in question can be solved using the models included in [15].

Other types of problems include transportation problems with minimal maintenance costs published in [16].

All the so far stated models belong among the so-called deterministic models. An independent group of the models is then formed by stochastic models of the transportation problems. Some of the first models were published in [8] and [9]. They are transportation problems with random transportation expenses, customer stochastic requirements and stochastic production – transportation problems. Considering the approaches used for solving transportation problems with a stochastic character, we can point out the approach that is based on fuzzy logic, published in [17] or solved in the conditions of fuzzy networks [18]. From among other stochastic variants of the transportation problems, we can also point out the fuzzy transportation problem solved in [19] or the fuzzy

transportation problem with integral flow published in [20]. Yet another type of a transportation problem affected by uncertainty is the so-called stochastic bottleneck transportation problem with flexible supply and demand quantity published in [21] or [22].

Nevertheless, we have not found any publication that would present a transportation problem variant with time windows. Time windows can be important for acceptability of proposed solutions. They particularly include time limitation cases, in which we are supposed to practically implement delivery of a commodity.

3 Problem Formulation and Mathematical Model Proposal

We have a set of sources I and a set of customers J. For each source $i \in I$ there is a defined capacity, time interval that represents the operation time $\langle t_i^{s1}; t_i^{s2} \rangle$, where t_i^{s1} is the beginning of the operation time (the earliest possible time loading of a given vehicle can commence) and t_i^{s2} is the end of the operation time (the latest possible time loading of a given vehicle can be completed), and loading time of the given commodity unit n_i. For each customer $j \in J$ there is a defined customer requirement, time interval that represents the operation time $\langle t_j^{c1}; t_j^{c2} \rangle$, where t_j^{c1} is the beginning of the operation time (the earliest possible time unloading of a given vehicle can commence) and t_j^{c2} is the end of the operation time (the latest possible time loading of a given vehicle can be completed), and loading time of the given commodity unit v_i. Only one time window is defined for each source and each customer (operation time without interruption during the planned period).

For each relation between a source $i \in I$ and customer $j \in J$ there is a defined cost for transporting the given commodity unit c_{ij} and transportation time T_{ij}. Moreover, the cost of vehicle downtime during the time unit p is known as well.

We also expect that it is not strictly required that the commodity is delivered by the beginning of the customer operation times, at the latest (which is the case when it comes to, for example, bakery products etc.). The vehicle arrival time to the customer can thus occur at any time within the customer's operation time. The only limitation is that unloading of the commodity has to be completed during the operation time. Analogically, we assume that no maximal number of vehicles that can be simultaneously loaded at the sources and unloaded at the customer facilities is defined.

We do not consider capacity limitations of individual vehicles for now. It means that a maximum of one shuttle takes place between the given source and customer. When the commodity is transported, one shuttle takes place. When the commodity is not transported, the shuttle does not take place.

The objective is to decide about the number of transported units between individual sources and customers in a way that the overall transportation cost is as little as possible.

For the purpose of modelling the necessary decisions, we will introduce the following variables to the problem:
x_{ij}… number of commodity units transported from source $i \in I$ to customer $j \in J$,
t_{ij}… departure time of the loaded vehicle from source $i \in I$ to customer $j \in J$,
τ_{ij}… waiting time of the vehicle arriving from source $i \in I$ to customer $j \in J$ until the commencement of its operation time.

When we want the departure times in the relations without transportation to acquire, for example, values $t_{ij} = 0$, we introduce a new group of auxiliary bivalent variables y_{ij} for $i \in I$ and $j \in J$ to the model. If, after the completion of the optimization calculation, $y_{ij} = 1$, then the transport between source $i \in I$ and customer $j \in J$ does not take place; otherwise $y_{ij} = 0$.

The mathematical model that operates with time constraints related to the given limited operation times of individual sources and customers, which minimizes the overall cost that consists of the overall transportation cost increased by the overall cost related to vehicle downtimes, can have the following form:

$$\min f(x, y, t, \tau) = \sum_{i \in I} \sum_{j \in J} \left(c_{ij} x_{ij} + p \tau_{ij} \right) \tag{1}$$

subject to:

$$\sum_{j \in J} x_{ij} = a_i \qquad \text{for } i \in I \tag{2}$$

$$\sum_{i \in I} x_{ij} = b_j \qquad \text{for j} \in \text{J} \qquad (3)$$

$$t_i^{s1} + n_i x_{ij} \leq t_{ij} + t_i^{s1} y_{ij} \qquad \text{for } i \in I \text{ and } j \in J \qquad (4)$$

$$t_{ij} + t_i^{s2} y_{ij} \leq t_i^{s2} \qquad \text{for } i \in I \text{ and } j \in J \qquad (5)$$

$$t_j^{c1} \leq t_{ij} + T_{ij}(1 - y_{ij}) + \tau_{ij} + t_i^{c1} y_{ij} \qquad \text{for } i \in I \text{ and } j \in J \qquad (6)$$

$$t_{ij} + T_{ij}(1 - y_{ij}) + \tau_{ij} + v_j x_{ij} + t_i^{c2} y_{ij} \leq t_j^{c2} \qquad \text{for } i \in I \text{ and } j \in J \qquad (7)$$

$$x_{ij} \leq M_{ij}(1 - y_{ij}) \qquad \text{for } i \in I \text{ and } j \in J \qquad (8)$$

$$1 - y_{ij} \leq x_{ij} \qquad \text{for } i \in I \text{ and } j \in J \qquad (9)$$

$$x_{ij} \geq 0 \qquad \text{for } i \in I \text{ and } j \in J \qquad (10)$$

$$t_{ij} \geq 0 \qquad \text{for } i \in I \text{ and } j \in J \qquad (11)$$

$$\tau_{ij} \geq 0 \qquad \text{for } i \in I \text{ and } j \in J \qquad (12)$$

$$y_{ij} \in \{0; 1\} \qquad \text{for } i \in I \text{ and } j \in J \qquad (13)$$

Function (1) represents an optimization criterion – overall transportation cost increased by the cost related to vehicle downtimes when waiting for unloading at customer facilities. Group of constraints (2) will ensure that the capacities of all sources will be exhausted. Group of constraints (3) will ensure compliance with the given requirements of all customers. Groups of constraints (4) and (5) ensure that loaded vehicles leave upon the loading completion, at the earliest, and by the end of the source operation time, at the latest. Groups of constraints (6) and (7) ensure that vehicle unloading takes place during the given customer operation time. Furthermore, group of constraints (6) will also allow for the calculation of possible vehicle downtime when waiting for unloading at the given customer facilities. Group of constraints (8) will ensure that when $x_{ij} > 0$, then $y_{ij} = 0$. Group of constraints (9) will ensure that when $x_{ij} = 0$, then $y_{ij} = 1$. Groups of constraints (10) – (13) specify the domains of definition of the variables included in the model. Values M_{ij} represent matrix elements of the prohibitive constants, calculated based on the following relations:

$$M_{ij} = min\{a_i; b_j\} \qquad \text{for } i \in I \text{ and } j \in J \qquad (14)$$

When creating the model, we restricted it to the transportation problem variant in a balanced form. For the other, unbalanced forms, the model can be easily adjusted by modifying groups of constraints (2) and (3).

4 Calculation Experiments with the Proposed Model

To demonstrate the proposed model, we chose the following model problem. Let us say there are 5 sources with their respective capacities and 5 consumers with their respective requirements – see Table 1.

Source	**1**	**2**	**3**	**4**	**5**
Source capacity	5	4	3	2	1
Source operation time	6:00 a.m. – 1:00 p.m.	6:00 a.m. – 1:00 p.m.	7:00 a.m. – 2:00 p.m.	7:00 a.m. – 2:00 p.m.	7:00 a.m. – 3:00 p.m.
Customer	1	2	3	4	5
Customer requirements	1	2	3	4	5
Customer operation time	8:00 a.m. – 8:00 p.m.	12:00 p.m. – 10:00 p.m.	6:00 a.m. – 10:00 p.m.	7:00 p.m. – 12:00 a.m.	2:00 p.m. – 11:00 p.m.

Table 1 Input data related to sources and customers

The values of individual unit transportation costs and transportation times between the sources and customers are stated in Table 2 and Table 3.

	Customer 1	**Customer 2**	**Customer 3**	**Customer 4**	**Customer 5**
Source 1	7	5	3	3	4
Source 2	2	10	7	6	1
Source 3	5	4	4	3	2
Source 4	0	9	7	4	10
Source 5	4	3	3	4	10

Table 2 Values of the unit transportation costs for the relations between the sources and the customers

	Customer 1	**Customer 2**	**Customer 3**	**Customer 4**	**Customer 5**
Source 1	10	8	5	6	13.5
Source 2	8	4	8	10	11
Source 3	12	6	4.5	13	18
Source 4	0	6	4	5.5	1
Source 5	6	7	4	8	5

Table 3 Values of the transportation times for the relations between the sources and the customers

The cost per a vehicle downtime time unit is considered to be $p = 5$ of the monetary units. The models were solved using the Xpress-IVE optimization software. The calculation time did not exceed 1 s. The result of the optimization calculation is represented by the following numbers of transported units x_{ij} (Table 4) and vehicle departure time t_{ij} from source $i = 1, \dots, 5$ to customer $j = 1, \dots, 5$ (Table 5).

	Customer 1	**Customer 2**	**Customer 3**	**Customer 4**	**Customer 5**
Source 1	0	0	3	1	1
Source 2	0	0	0	0	4
Source 3	0	1	0	2	0
Source 4	1	0	0	1	0
Source 5	0	1	0	0	0

Table 4 Final values of variables x_{ij}

	Customer 1	**Customer 2**	**Customer 3**	**Customer 4**	**Customer 5**
Source 1	0	0	6:00	13:00	6:00
Source 2	0	0	0	0	6:00
Source 3	0	7:00	0	7:00	0
Source 4	8:00	0	0	13:30	0
Source 5	0	7:00	0	0	0

Table 5 Final values of variables t_{ij}

When applying this solution, it applied for the values of variables τ_{ij} ($i = 1, \dots, 5$ and $j = 1, \dots, 5$) that $\tau_{ij} = 0$. It means that, from the time perspective, the individual trips are planned in a way that no waiting and subsequent downtimes occur at the given customer facilities.

5 Conclusions

Transportation problems belong among the basic decision-making tasks in economic applications of mathematic modelling. The most common problem is the so-called classic transportation problem, in which real limitations are exclusively represented by source capacities and customer requirements. However, usability of the classic transportation problems for solving distribution issues in logistic systems is currently very limited. The main reason of this limitation is the fact that classic transportation problems are not able to take into account other real limitations, which normally occur in real life. In order to solve a realistic problem that shows signs of a

transportation problem with additional constraints, we can use the classic transportation problem, however, we need to subject the acquired results to an additional inspection that focuses on the fulfilment of the other limitations, which were not considered in the model. It can also happen that some of the additional limitations will not be fulfilled, thus essentially making the acquired solution inacceptable. In that case it has to be "manually changed", however, without a guarantee that the modified solution will be optimal.

Our paper also points out at another, so far unpublished type of a transportation problem, in which the sources and customers have time windows, in which they can issue and accept individual shipments. We have proposed a linear mathematical mode that allows for solving this problem in an exact way without fundamentally increasing the calculation time.

References

[1] Hitchcock, F. L. (1941). The Distribution of a Product from Several Sources to Numerous Localities. *Journal of Mathematics and Physics,* 20 (1-4), 224-230.

[2] Dantzig, G. B. (1951). Application of the Simplex Method to a Transportation Problem. In *Activity Analysis of Production and Allocation (T.C. Koopmans, ed.)* (pp. 359-373). New York: Wiley.

[3] Balinski, M. L. (1961). Fixed-cost transportation problems. *Naval Research Logistics Quarterly,* 8 (1), 41-54.

[4] Gottlieb, J. & Paulmann, L. (1998). Genetic algorithms for the fixed charge transportation problem. In *IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence* (pp. 330-335). Anchorage: *IEEE*.

[5] Sun, M., Aronson, J. E., McKeown, P. G. & Drinka, D. (1998). A tabu search heuristic procedure for the fixed charge transportation problem. *European Journal of Operational Research,* 106 (2-3), 441-456.

[6] Zhao, Y., Larsson, T., Rönnberg, E. & Pardalos, P. M. (2018). The fixed charge transportation problem: a strong formulation based on Lagrangian decomposition and column generation. *Journal of Global Optimization,* 72 (3), 517-538.

[7] Goldštejn, J. G. & Judin, D. B. (1969). *Zadači linejnogo programmirovanija transportnogo tipa.* Moskva: Nauka.

[8] Korda, B. (1967). *Matematické modely v ekonomii.* Praha: SNTL.

[9] Walterová, L. (1986). *Dopravní modely.* Státní pedagogické nakladatelství, Praha.

[10] Janáček, J. (1999). *Matematické programování.* Žilina: Žilinská univerzita.

[11] Palúch, S. & Peško, Š. (2006). *Kvantitatívne metódy v logistike.* Žilina: EDIS.

[12] Teichmann, D., Ivan, M. & Grosso, A. (2013). Matematické modely špecifických dopravných úloh. *Železničná doprava a logistika,* IX (1), 12-17.

[13] Teichmann, D., Dorda, M. & Mocková, D. (2017). Transportation Problem Model Supplemented with Optimisation of Vehicle Deadheading and Single Depot Parking. In *Proceedings of 35th International Conference Mathematical Methods in Economics (MME): 35th international conference* (pp. 789-794). Hradec Králové: Gaudeamus.

[14] Cenek, P. & Jánošíková, Ľ. (2008). *Models and Optimization in Transports and Logistics.* Žilina: EDIS.

[15] Graf, V., Teichmann, D. & Dorda, M. (2017). Transportation Problem with degressively stepped costs. In *Proceedings of Mathematical Methods in Economics MME 2017: 35th international conference* (pp. 201-206). Hradec Králové: Gaudeamus.

[16] Kibzun, A. I. & Khromova, O. M. (2016). Mathematical Modelling of a Transport System with Minimal Maintenance Costs. *Bulletin of the South Ural State University. Series "mathematical Modelling, Programming and Computer Software",* 9 (3), 41-54.

[17] Chanas, S., Kołodziejczyk, W. & Machaj, A. (1984). A fuzzy approach to the transportation problem. *Fuzzy Sets and Systems,* 13 (3), 211-221.

[18] Chen, M., Ishii, H. & Wu, C. (2008). Transportation problems on a fuzzy network. *International Journal of Innovative Computing Information and Control,* 4 (5), 1105-1109.

[19] Ishii, H., Tada, M. & Nishida, T. (1990). Fuzzy Transportation Problem. *Journal of Japan Society for Fuzzy Theory and Systems,* 2 (1), 79-84.

[20] Tada, M. & Ishii, H. (1996). An integer fuzzy transportation problem. *Computers & Mathematics with Applications,* 31 (9), 71-87.

[21] Geetha, S. & Nair, K. P. K. (2017). A Stochastic Bottleneck Transportation Problem. *Journal of the Operational Research Society,* 45 (5), 583-588.

[22] Ge, Y. & Ishii, H. (2011). Stochastic bottleneck transportation problem with flexible supply and demand quantity. *Kybernetika,* 47 (4), 560 – 571.

Task threatness matrix in the Project management

Brožová Helena[1], Rydval Jan[1], Pavlíčková Petra[1], Šubrt Tomáš[1],

Abstract. The main aim of the project management is to ensure the successful completion of a project which consists of many tasks. The success of a project depends on many constraints as scope, time, quality, budget and allocation of necessary recourses. The crucial question is to find out the tasks which can cause the delaying of a project time or failure of the objectives of project. To find out these tasks is necessary to analyze all project tasks and to determine their threatness in relation to the project objectives.

In this paper we analyze project task using the criticalness potential defining the criticalness of tasks and using the failureness potential defining the possibility of various tasks failure. Criticalness potential of task has broader meaning then critical task, it means evaluation of time, topological, slack, cost and work criticalness of task using multiple decision-making methods. Failureness of tasks is analyzed using fuzzy linguistic terms and the proper fuzzy linguistic scale. The aggregation of the evaluation of these two potentials expresses the tasks threatness assessment which can be displayed in the task threatness matrix.

Keywords: Project management, task, criticalness potential, failureness potential, task threatness, threatness matrix.

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

1 Introduction

Project management is a scientific and managerial discipline dealing with all phases of project live cycle as initiating, planning, executing, controlling, and closing with the aim of managing a project team to achieve all project goals described by various specific project criteria as effectively as possible. The principal purpose of project management is a successful implementation of a project: finishing the project within planned time, with planned costs and with a fulfilled objective. The aim of the project management is to fulfil the project management triangle ([11], [15], [12]). The success of the project is conditioned by the level of recognition of weak spots in the project. The criticalness of project activities, such as the level of amount and strength of weak spots in the project, is not only given by the surroundings and environment of the project but also by internal arrangement and structure of the project ([3]).

In the area of risk management, a number of different techniques and approaches have been derived in order to reduce the risk of a project and its partial tasks ([16]). The criticality of project tasks is often defined from a time perspective only, using stochastic approaches ([1]) and ([8]), fuzzy sets methods ([7], [19]) or using the findings of a network analysis ([9]) and ([3]). Gong and Rowings ([9]) mention that ignoring the impact of non-critical tasks, which may easily become critical, is the most frequent criticism of project duration analysis methods. Another point of view on tasks criticalness is given by the structure of relations in the project. Bowers ([1]) or Williams ([17]) deal with a stochastic analysis of a project network where the criticality of tasks in the project is derived from the relation between task duration and the whole project, and on the basis of a number of resources used for a task and the whole project.

Quantitative estimation serves for the evaluation of the task criticalness. Although such quantitative estimation of the project is performed in a pre-project phase, generally only traditional time and resource analysis of the project is provided. Primarily, the activities are evaluated from a time perspective only as critical or uncritical activities, but many other quantitative characteristics can be used. Such a multiple attribute analysis enables not to ignore the impact of time-uncritical activities which may have the high criticalness from many different reasons ([3]). Our point of view is that, the criticalness, the failureness and the threatness of the project tasks should be evaluated and managed. Because if they are not leaded the project could fail in all aspects of the project management triangle.

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management,
Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, brozova@pef.czu.cz

The aim of this paper is to analyze project task threatness using the criticalness potential defining the criticalness of tasks and using the failureness potential defining the possibility of various tasks failure. The aggregation of the evaluation of these two potentials expresses the tasks threatness assessment which can be displayed in the task threatness matrix.

2 Task threatness and task threatness matrix

Task threatness is defined as two-dimensional task evaluation aggregating multiple attributes. All attributes are divided into two groups. The first one is a group consisting of five indicators allowing crisp evaluation based on the natural units as time, cost, amounts of resources etc. The second one consists of three indicators which have to be evaluated using soft tools, using fuzzy linguistic scale in this case. The final task threatness is received as two-dimensional evaluation based on aggregated criticalness potential and failureness potential. Structure of these attributes is in the Figure 1.

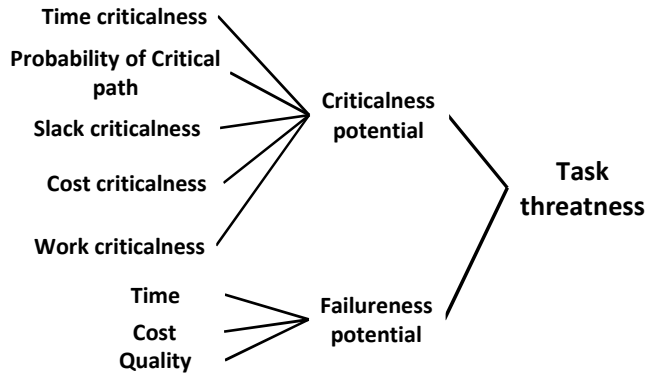


Figure 1 Hierarchical structure of task threatness

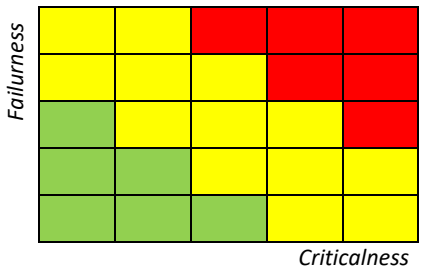


Figure 2 Structure of task threatness matrix

Suggested two-dimensional evaluation can be expressed in the Winterling matrix which is called the task threatness matrix ([18]). We suggest the first axis describing the criticalness potential, the second one describing the failureness potential of the tasks, and five levels of both criticalness and failureness potentials. The tasks are placed into the matrix cells based on both potentials evaluations. The more task moves diagonally up right, the more attention should be given to it within the project management. The red area consists of the tasks (red task) which require intensive supervision to ensure successful completion of the project. The yellow tasks should be monitored and controlled, because these tasks could negatively influence success of the project. In the green area, there are tasks which do not represent serious problem for project success.

2.1 Indicators of task criticalness and criticalness potential

Brozova et al. ([2], [3], [4]) suggested to provide the overall evaluation of the task criticalness using quantitative crisp evaluation without soft knowledge about character of the tasks. The criticalness potential of the project tasks is based on the multiple attribute decision making method using five indicators of the criticalness as time duration, time slack, cost, and work and probability that the task will lie on critical path.

This probability that some (maybe critical) path gateway through task is calculated as

$$p_0 = 1 \quad \text{and} \quad p_i = \sum_{j \ predeces \ i} \frac{p_j}{h_j}, \ i = 1,2,\dots,N \tag{1}$$

where p_i are the probabilities that the task i will be on the critical path, and h_j is the number of tasks following task j, (task 0 is the unique starting task frequently dummy), N is the number of the other tasks in the project. Other indicators are calculated during project planning.

If the project manager is risk neutral, the values of these indicators are proportionally converted into interval $\langle 1,10 \rangle$ using formula

$$u_i = 9 \frac{\min\limits_k v_k - v_i}{\min\limits_k v_k - \max\limits_k v_k} + 1 \tag{2}$$

where v_i, v_k are criticalness evaluations and u_i is normalized value. The value 1 corresponds to the lowest partial criticalness and the value 10 to the highest criticalness.

This type of normalization is used because it allows to use the multiplicative model of criticalness indicators aggregation ([10]) which emphasizes task criticality (risk averse position)

$$CP_i = \prod_n u_{in} \tag{3}$$

where CP_i is criticalness potential of the task i, u_{in} is evaluation of each indicators of the criticalness and n is number of used indicators of criticalness. The task criticalness potential is in this case calculated using formula

$$CP_i = \left(9\frac{p_i - \min\limits_k p_k}{\max\limits_k p_k - \min\limits_k p_k} + 1\right) \times \left(9\frac{t_i - \min\limits_k t_k}{\max\limits_k t_k - \min\limits_k t_k} + 1\right) \times \left(9\frac{s_i - \max\limits_k s_k}{\min\limits_k s_k - \max\limits_k s_k} + 1\right) \times$$
$$\times \left(9\frac{c_i - \min\limits_k c_k}{\max\limits_k c_k - \min\limits_k c_k} + 1\right) \times \left(9\frac{w_i - \min\limits_k w_k}{\max\limits_k w_k - \min\limits_k w_k} + 1\right) \tag{4}$$

where CP_i is the criticalness potential of the task i , p_i is the topological location which reflects the probability the critical path will pass through the current task i, t_i is the duration of the task i, s_i is the time slack of the task i, c_i is the cost of the task i, w_i is the work amount of the task i.

2.2 Indicators of task failureness and failureness potential

The soft evaluation of the project tasks based on experts' knowledge and experts' experience of characters of tasks represents the task failureness ([2]). Task failureness potential is based on the project triangle view with three aspects: cost, time and quality failureness. The evaluation method is based on the soft system approach ([6]) because many factors influencing the tasks success or failureness are based on uncertainty, probability and possibility or have social or human roots.

Suggested evaluation of the task failureness indicators uses linguistic terms and fuzzy scales ([13], [14]). The non-uniform six-point fuzzy scale here used allows realistic subjective evaluation of the task failureness where uncertainty decreases towards the extreme values. We consider trapezoid membership functions described in Figure 3whose representation is achieved by 4-tuples $A = (a_1, a_2, a_3, a_4)$.

Task can be (timely, costly or qualitatively)	
Not at all failing	(0; 0; 0; 0.1)
Usually not failing	(0; 0.1; 0.2; 0.3)
Rather not failing	(0.2; 0.3; 0.4; 0.6)
Rather failing	(0.4; 0.6; 0.7; 0.8)
Usually failing	(0.7; 0.8; 0.9; 1)
Always failing	(0.9; 1; 1; 1)

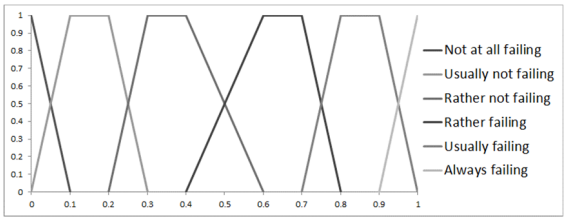


Figure 3. Linguistic fuzzy scale for task failureness evaluation

For the failureness potential evaluation is used a non-uniform five-point fuzzy scale with uncertainty increases towards the evaluation Extremely failureness task which is suitable for a risk-averse appetite of the project manager or owners of the project (Figure 4). The application of the suggested fuzzy system enables to obtain an approximative overall evaluation of the task failureness for all project tasks.

Task failureness potential	
Non-failing task	(0; 0; 0.05; 0.15)
Weakly failing task	(0.05; 0.15; 0.25; 0.35)
Rather failing task	(0.25; 0.35; 0.5; 0.6)
Strongly failing task	(0.5; 0.6; 0.75; 0.85)
Extremely failing task	(0.75; 0.85; 1; 1)

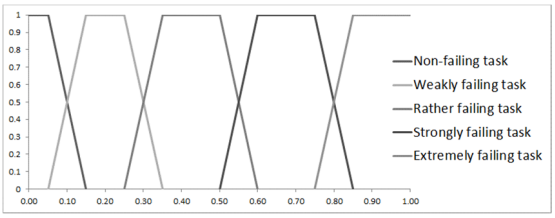


Figure 4. Linguistic fuzzy scale for task failureness potential

The failureness potential of the task aggregates all partial evaluations of the tasks failureness. The simplest form of aggregation uses weighted fuzzy sum

$$FP_i = (f_{i1}, f_{i2}, f_{i3}, f_{i4}) = \left(\left(\frac{a_{i1}+b_{i1}+c_{i1}}{3}\right), \left(\frac{a_{i2}+b_{i2}+c_{i2}}{3}\right), \left(\frac{a_{i3}+b_{i3}+c_{i3}}{3}\right), \left(\frac{a_{i4}+b_{i4}+c_{i4}}{3}\right)\right) \tag{5}$$

where $FP_i = (f_{i1}, f_{i2}, f_{i3}, f_{i4})$ is the failureness potential of the task i, $A_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4}), B_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})$, $C_i = (c_{i1}, c_{i2}, c_{i3}, c_{i4})$ are the fuzzy evaluations of the task i failureness from the project triangle criteria. This fuzzy evaluation is then compared with the linguistic fuzzy scale and task failureness potential is estimated using the proper principle of linguistic approximation.

3 Example of the tasks threatness matrix

The analysis of project tasks threatness is described on the following small-scale project ([2], [3], [4],). A critical path of a project (Figure 5) is composed of tasks A, C, D, F, I, J, K and N.

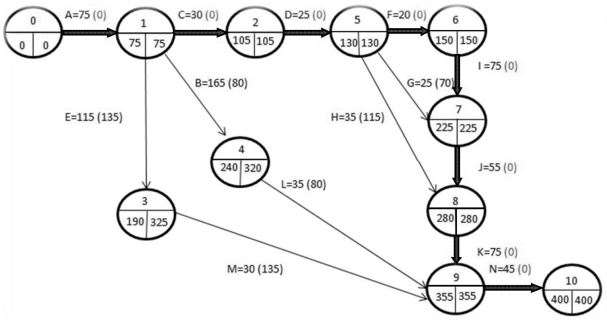


Figure 5 Small-scale project with the indication of a critical path

Individual criticalness factors give an information on how project task potentially endangers a successful realisation of a project from a different point of view. Estimation of the criticalness potential of the project activities is based on the multiple criteria decision-making methods using five indicators of the criticalness (Table 1).

Task	Days	Probability of critical path	Slack	Cost	Work	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Criticalness potential
A	75	1	0	1350	450	4.414	10.000	10.000	6.908	2.786	8495.05
B	165	0.33	80	990	495	10.000	3.225	4.667	5.260	2.992	2368.73
C	30	0.33	0	630	630	1.621	3.225	10.000	3.611	3.611	681.66
D	25	0.33	0	175	175	1.310	3.225	10.000	1.527	1.527	98.51
E	115	0.33	135	690	345	6.897	3.225	1.000	3.885	2.305	199.18
F	20	0.11	0	60	60	1.000	1.000	10.000	1.000	1.000	10.00
G	25	0.11	70	150	150	1.310	1.000	5.333	1.412	1.412	13.93
H	35	0.11	115	210	210	1.931	1.000	2.333	1.687	1.687	12.82
I	75	0.11	0	1575	1575	4.414	1.000	10.000	7.939	7.939	2782.04
J	55	0.22	0	1155	1155	3.172	2.112	10.000	6.015	6.015	2423.81
K	75	0.33	0	2025	2025	4.414	3.225	10.000	10.000	10.000	14235.15
L	35	0.33	80	875	875	1.931	3.225	4.667	4.733	4.733	651.06
M	30	0.33	135	360	180	1.621	3.225	1.000	2.374	1.550	19.24
N	45	1	0	945	945	2.552	10.000	10.000	5.053	5.053	6515.97

Critical path tasks are grey highlighted

Table 1 Criticalness potential of project tasks

The Figure 6 contains graphical description of values and differences between values of criticalness potential. From this graph is seen that 10 tasks have very low criticalness potential. Only one task has the highest criticalness potential. The real values of criticalness indicators, their 1-10 normalization and their criticalness potential using multiplicative aggregation are in the Table 1.

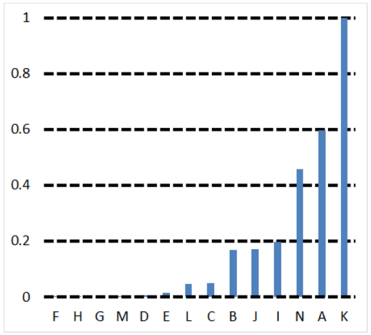


Figure 6 Ranked tasks according to their criticalness potential

Table 2 contains linguistic evaluation of the project tasks failureness together with aggregation of this evaluation using weighted sum method and their transformation into the task failureness potential. Figure 7 shows

the fuzzy expression of failureness potential and its comparisons with linguistic system. It is possible to see, that no task is extremely failing.

Task	Time failureness	Quality failureness	Cost failureness	Failureness potential				
A	usually not failing	rather failing	not at all failing	0.13	0.23	0.30	0.40	Weakly failing task
B	usually not failing	usually not failing	usually not failing	0.00	0.10	0.20	0.30	Weakly failing task
C	rather not failing	rather not failing	rather not failing	0.20	0.30	0.40	0.60	Rather failing task
D	usually not failing	usually not failing	usually not failing	0.00	0.10	0.20	0.30	Weakly failing task
E	usually failing	usually failing	usually not failing	0.47	0.57	0.67	0.77	Strongly failing task
F	usually not failing	not at all failing	not at all failing	0.00	0.03	0.07	0.17	Non-failing task
G	usually not failing	not at all failing	not at all failing	0.00	0.03	0.07	0.17	Non-failing task
H	usually not failing	not at all failing	not at all failing	0.00	0.03	0.07	0.17	Non-failing task
I	rather not failing	usually not failing	rather not failing	0.30	0.40	0.50	0.63	Rather failing task
J	rather failing	rather failing	rather failing	0.40	0.60	0.70	0.80	Strongly failing task
K	rather failing	rather failing	rather failing	0.40	0.60	0.70	0.80	Strongly failing task
L	usually failing	rather failing	rather not failing	0.43	0.57	0.67	0.80	Strongly failing task
M	usually not failing	not at all failing	not at all failing	0.00	0.03	0.07	0.17	Non-failing task
N	rather not failing	rather not failing	rather not failing	0.20	0.30	0.40	0.60	Rather failing task

Critical path tasks are grey highlighted

Table 2 Failureness potential of project tasks

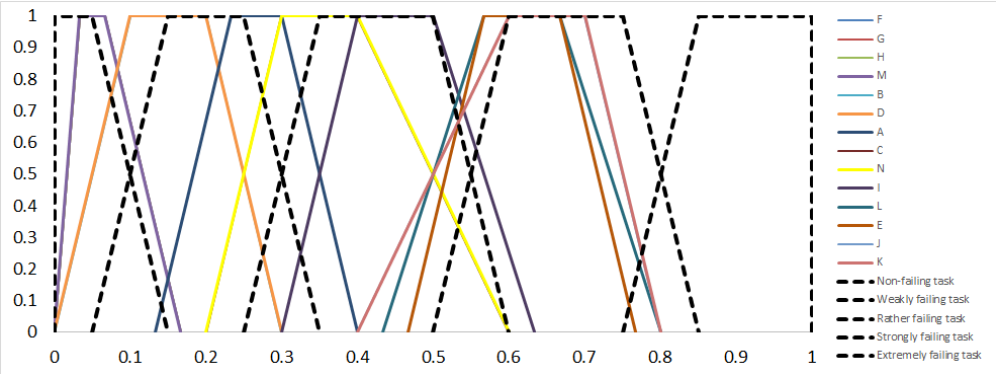
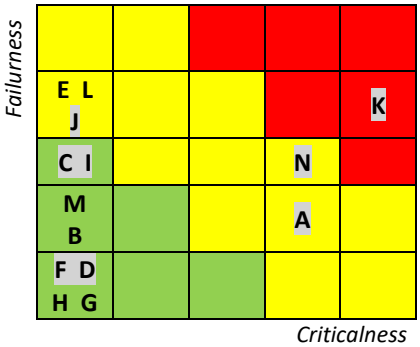


Figure 7 Ranked tasks according to their failureness potential

The values of criticalness and failureness potentials can be divided into five levels and the task threatness matrix is created (Figure 8). In the red area of task threatness matrix, there is only one highly threatening task requiring great attention (K), and in the yellow area there are five threatening tasks to be controlled, to ensure the successful completion of the project (E, L, J, N, A). Remaining eight tasks in green area (C, I, M, B, F, D, H, G) should not significantly influence the project.



Critical path tasks are grey highlighted

Figure 8 Task threatness matrix for small project

Using the linguistic terms and fuzzy scales to evaluating the task failureness indicators is a suitable way how to express the individual and subjective opinion on the task failureness possibilities ([13], [14]) but it should be pointed out, that results are affected by the type of fuzzy scale (i.e. uniform or non-uniform).

Furthermore, the suggested two-dimensional evaluation of task threatness should be adapted to the individual project environment and project needs. It is necessary to set up the proper scale size (for example three, five of seven points), and to divide the matrix into the areas, according to the selected scale size ([18]). Of course, the size of each zone can also be individually tailored to specific needs of the project management.

4 Conclusion

This paper suggests how to analyze project task threatness using the criticalness potential defining the criticalness of tasks and using the failureness potential defining the possibility of various tasks failure. Two-dimensional evaluation of task threatness is displayed in the task threatness matrix. This two-dimensional assessment of the project tasks is useful with respect to the project management triangle, because it involves a multi-criteria assessment of the tasks and their impact on the project success. The evaluation process described above can be the basis of a module in project management software as a support for project managers' decision-making showing project weaknesses in advance. Important advantage of suggested threatness matrix is that allows to combine sharp and fuzzy assessments of the impact of individual activities on project completion.

Acknowledgements

The research is supported by the Operational Programme Prague – Growth Pole of the Czech Republic - Implementation proof-of-concept activities CULS to promote technology transfer and knowledge into practice. number: CZ.07.1.02/0.0/0.0/17_049/0000815 - KZ10.

References

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

[2] Brožová, H., Bartoška, J. & Šubrt, T. (2014). Fuzzy Approach to Risk Appetite in Project Management. *Proceedings of the 32st International conference on Mathematical Methods in Economics* (pp.61–66). Olomouc: Palacký University.

[3] Brožová, H., Bartoška, J., Šubrt, T. & Rydval, J. (2016). Task Criticalness Potential: A Multiple Criteria Approach to Project Management. *Kybernetika*, *52*, 558-574.

[4] Brožová, H. & Rydval, J. (2017). Modelling synergy of the complexity and criticalness factors in the Project management. *Proceedings of the 35th International Conference on Mathematical Methods in Economics* (pp. 84-89). Hradec Králové: University of Hradec Králové.

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

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

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

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

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

[10] Hwang, Ch. L. & Yoon, K. (1981). *Multiple Attribute Decision Making*. Berlin, Heidelberg, New York: Springer Verlag.

[11] Kerzner, H. (1979). *Project Management: A Systems Approach to Planning, Scheduling and Controlling*. New York: Van Nostrand Reinhold.

[12] PMI (2013). *A Guide to the Project Management Body of Knowledge*, 5th ed., Newtown Square: Project Management Institute.

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

[14] Wang, Z. & Klir, G. J. (1992). *Fuzzy Measure Theory*. New York: Plenum Press.

[15] Warburton, R. D. H. (2011), "A time-dependent earned value model for software projects", *International Journal of Project Management*, *29*, 1082-1090.

[16] Wideman, R. M. (1992). *Project and Program Risk Management*. Newtown Square: Project Management Institute.

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

[18] Winterling, K. (1989). Risiken in Unternehmen erkennen, verhindern, bewältigen, *Management Zeitschrift*, *58*, 30-33.

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

Analyzing of mortality of enterprises created under the Human Capital Operational Programme - a case study

Anna Szmit[1], Maciej Szmit[2], Dominika Lisiak-Felicka[3]

Abstract. However the original laws of mortality were formulated to human population, the researches of the dynamics of mortality (or – which is the same on the other side – the dynamics of survivability) are nowadays conducted in various fields, starting from medicine, through actuarial science up to economy and managerial sciences. The subject of the research, the results of which will be presented in the article, is the dynamics of mortality of over 350 microenterprises created in Lodz region (Poland) during 2007-2013 years under the Human Capital Operational Programme, in a few projects under the Measure 6.2 and the Sub-measure 8.1.2, in the years 2007-2013, with the support of the Fundacja Inkubator from Lodz. The article is a continuation and extension of our previous research including three well-known models (Exponential, Weibull and Log-logistic) and three advanced models proposed by Beard, Wilkie and Barnett.

Keywords: Survival Analysis, Beard Model, Barnett Model, Wilkie Model, Enterprises' Survival.

JEL Classification: M13
AMS Classification: 62M10

1 Introduction and literature review

The phenomenon of mortality and its statistical characteristics is considered in several disciplines such as demography (see e.g. [13], [29]), actuarial sciences (see e.g. [10], [8]) and population ecology (see e.g. [18], [26]). The original laws of mortality (a mathematical formulas describing how the mortality is depending on age) were formulated to human population. A typical pattern of mortality consist of three periods: children age (infancy with relatively high mortality and a rapid decrease of mortality during the first few years of life), second period where the deaths are mainly due to accidents and the third period is the almost geometric increase of mortality with age (and increase slowing down after age 80, so-called late life mortality deceleration (see e.g. [13], [10]).

Usability of such models to other entities (such as organizations, including enterprises) may be controversial because of other conditions affecting the functioning of economic life and characteristics of particular enterprise (see e.g. [12], [9], [15], [17], [22], [25]) or entrepreneur (see e.g. [6], [11], [19], [20], [24], [30]) or the possibility of far-reaching transformations such as ownership changes, changes in the form of business activity, splits and mergers etc. Additionally considering the survival of enterprises in various economies, it is needed to consider a number of factors, such as the economic condition, current economic policy, and even cultural determinants in particular countries (see e.g. [1], [2]).

In the context of the research potentially low survivability level may indicate negative effect of "forced entrepreneurship" consisting of enrolling to the Human Capital Operational Programme (HCOP) supported project people with lower than average chances of finding a job, who have small or even no competences to became successful entrepreneurs, but high survivability level may indicate to the occurrence of „creaming effect" implied by – unwanted from HCOP point of view – participation of persons, who would decide for becoming entrepreneurs even without HCOP support (see e.g. [4]). The aim of our research is to find the best way to describe microenterprises mortality phenomenon. The article is a continuation and extension of our previous research covering lifecycles of over 500 microenterprises created in Lodz region (Poland) during 2007-2013. In our previous works [28], [27] we analyzed fitting of a few elementary models: Naïve, Gompertz–Makeham [14], [23] and special variant of Pareto type II distribution i.e. Lomax distribution [21] and generalization of Pareto distribution i.e. Burr type XII distribution [7], [3] and – additionally – Holt smoothing method [16] as sample model containing an autoregressive component, due to previously identified difficulties in adjusting functions differentiable with respect to time variable, to the empirical data. Additionally we consider modified Gompertz-

[1] Lodz University of Technology, Department of Management, anna.szmit@p.lodz.pl.
[2] University of Lodz, Department of Computer Science, maciej.szmit@uni.lodz.pl.
[3] University of Lodz, Department of Computer Science in Economics, dominika.lisiak@uni.lodz.pl.

Makeham model (without assumption concerning increase of mortality with age). In the article fitting of a few other models: Weibull, Log-logistic, Beard model [5], Wilkie and Barnett, base on [10], have been compared.

2 Subject of study

Fundacja Inkubator (FI) is a non-profit organization located in Lodz. Its main goal is helping in starting microenterprises and promotion of entrepreneurship. A larger part of such enterprises is created with support of European Union programs, especially under the Human Capital Operational Programme.

In our research we analyzed microenterprises created in a few projects under the Measure 6.2 (Support and promotion of entrepreneurship and self-employment) and under the Submeasure 8.1.2 (Support to adaptation and modernization processes in the region) in the years 2007-2013 in the Lodz Region. Within these projects more than 500 person from Lodz region received grants to start their own businesses. Because one of the conditions for the newly created enterprises was to remain alive for the first 12 months (the survivability is forced to be 100%) we modeled the period starting from 13th month of each enterprise's life. Another issue concerns step changes in the time series after the first two years of enterprise's life. This situation is caused mostly by some tax privileges existing in Polish law system for start-ups during their first two years of live. Additionally we modeled the survivability in the 5,5 years' time horizon so only the enterprises started at least from 5,5 years ago were selected.

The detailed description of the investigated population is included in the book [28]. The investigated time series was aggregated to 1-month resolution. The survival rates of newly created microenterprises for ones existing for at least 3, 4, 5,5 and 6 years are presented on Figure 1.

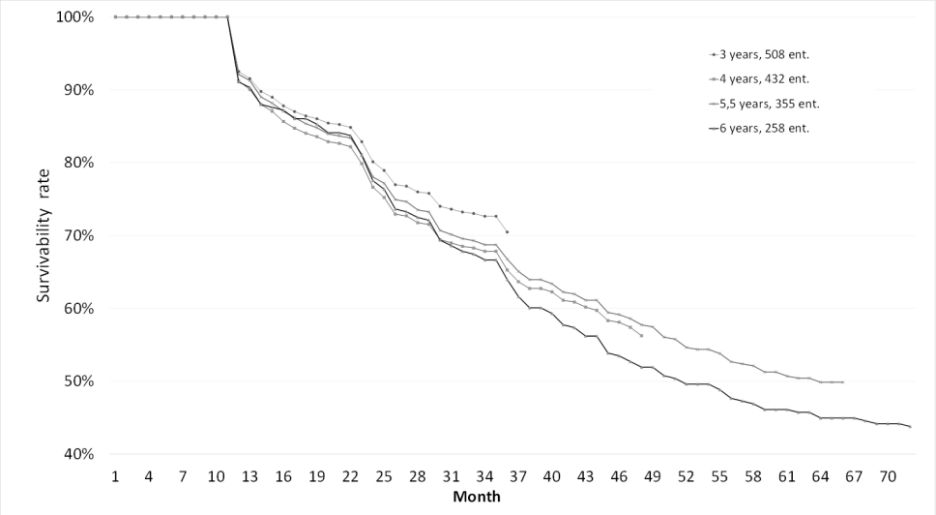


Figure 1 A comparison of survivability of enterprises created under the Human Capital Operational Programme in the Fundacja Inkubator. Source: [27]

3 Methodology

As described above we compared of fitting of several models to the empirical data. In the article fitting of two classics and three less widely used models of enterprises survivability are compared. Survival functions in these models are given respectively by formulas:

Weibull model:

$$S(t) = exp(-\alpha t^c) \tag{1}$$

Log-logistic:

$$S(t) = \frac{b}{b+t^c} \tag{2}$$

where:

- $S(t)$ – survivability (percent of enterprises remaining live) in month $t,\ t = 1, \dots, n$,
- b, c, α – parameters.

We also used model based on formula proposed by Beard, who analyzed the table of assured lives in the United Kingdom. However the original formula described mortality, it can simply be transformed to calculate survivability:

$$S(t) = 1 - q_t = 1 - \left(A + \frac{Bc^t}{Ec^{-2t} + 1 + Dc^t}\right) \tag{3}$$

where:

- q_t – mortality (percent of enterprises dead) in month $t,\ t = 1, \ldots, n$,
- A, B, c, D, E – parameters.

Two next models belong to the model's family based on general form given by formula:

$$S(t) = 1 - q_t = \frac{1}{1 + f(t)} \tag{4}$$

where:

- $f(t)$ – model-specific function.

The simplest model from this family is Barnett model, where:

$$f(t) = A - Ht + Bc^t \tag{5}$$

where:

- A, H, B, c – parameters.

The formula was used in the UK to graduate the experience of insured lives over the period 1967-1970 [10].

The other model was proposed by Wilkie, where function f(t) is a sum of components given by formula (6) and may contain various number s of parameters α_i:

$$f(t) = \exp\left(\sum_{i=1}^{s} \alpha_i t^{i-1}\right) \tag{6}$$

Best results in the described case were achieved for s=3, i.e.:

$$f(t) = exp(\alpha_1 + \alpha_2 t + \alpha_3 t^2) \tag{7}$$

Formula (4) is used on a few other models, e.g. First and Second Heligman-Pollard models (see [10]), using functions with greater number of parameters, but in our case the relatively small number of observations implied impossibility estimation of too big number of parameters in too complicated functions.

The above's models were compared with several ones considered in the article [27] i.e.: Gompertz-Makeham model with survival function given by equation:

$$S(t) = exp\left(\frac{B}{\ln c}(1 - c^t) - At\right) \tag{8}$$

where:

- A, B and c – parameters.

Burr Type XII distribution with survival function given by equation:

$$S(t) = \left(\frac{b}{b + t^c}\right)^s \tag{9}$$

where:

- $b > 0, c > 0, s > 0$ are parameters.

Pareto model, where distribution's survival function is a special case of the Burr Type XII distribution (c=1), given by the equation:

$$\mathrm{S(t)} = \left(1 + \frac{t}{b}\right)^{-s} \tag{10}$$

Holt exponential smoothing model (double exponential smoothing) given by equations:

$$L_t = \alpha y_{t-1} + (1 - \alpha)(L_t - 1 + T_{t-1}) \tag{11}$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \tag{12}$$

$$S(t) = L_t + T_t \tag{13}$$

where:

- L_t – the smoothed value for time t,
- T_t – estimate of the trend at time t,
- α – the data smoothing factor, $0 < \alpha < 1$, and
- β – the trend smoothing factor, $0 < \beta < 1$.

Because of obligatory selection of initial values in the Holt model, it's matching for the first observations may not be adequate, therefore the period from t = 15 was assumed for determining the average error. For ensuring comparability the same rule was applied for all other models.

All models' parameters were estimated using non-linear least squares method with Generalized reduced gradient (GRG) algorithm. Among the characteristics of the models' residuals distributions, the estimation of autocorrelation by first-order residual autocorrelation Pearson coefficient was presented.

As a measure of fitting we used Root Mean Square Percentage Error, RMSPE:

$$RMSPE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{S(t)-\hat{S}(t)}{S(t)}\right)^2} \tag{14}$$

where:

- $S(t)$ – original survivability value in month $t,\ t = 1, \dots, n$,
- $\hat{S}(t)$ – value of survivability estimated by considered model.

The mentioned characteristics of the data (including step changes) implicate potential issues concerning models' residual autocorrelation. Due of that, the first-order residual autocorrelation (given by Pearson correlation coefficient) was analyzed.

4 Results and conclusions

The measure of fitting of the results of all models to the original data (RMSPE) with information about number of estimated parameters is presented in Figure 2. The calculation was made for months 15-66 after enterprise started. Additionally, the first-order residual autocorrelation coefficient (r_e) is included. The arrow means that the model above (at the end of the arrow) is a generalization of the model below, cf. [24].

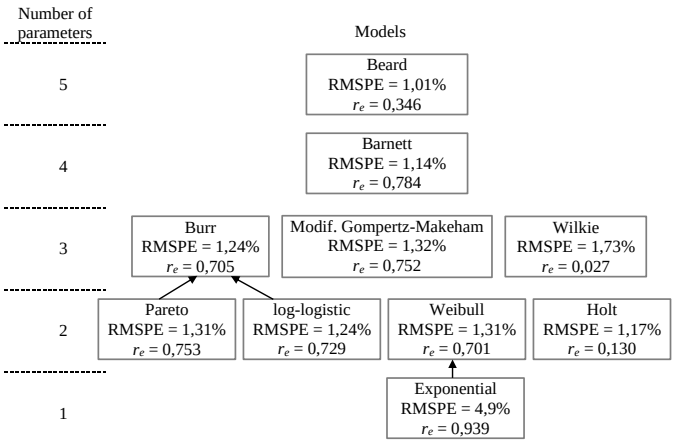


Figure 2 A comparison of survivability of enterprises created under the Human Capital Operational Programme in the Fundacja Inkubator. Source: own research

The original data and value of survivability estimated by above models are presented in Figure 3.

The fitting of all models is relatively high. The matching errors of the Beard and Barnett models are smaller than errors in the previously used models, whereas for the Wilkie the model did not provide an acceptable or satisfactory fit to the empirical data. As might be expected as the number of parameters increases, the accuracy of matching is almost always improved (excluding Holt model).

It is also noteworthy that the autocorrelation of residues in the Beard model is definitely smaller than in the previously used models (apart from the Holt adaptive model) while the autocorrelation in Barnett model is relatively high (however it is still smaller than autocorrelation in simplest models – Burr, Pareto and Modified Gompertz-Makeham).This may indicate the complexity of the dynamics of the Small and Medium Enterprises

dying process and the need to investigate the other, even more complex models, however the relatively small number of observations implied impossibility estimation of parameters of too complicated functions.

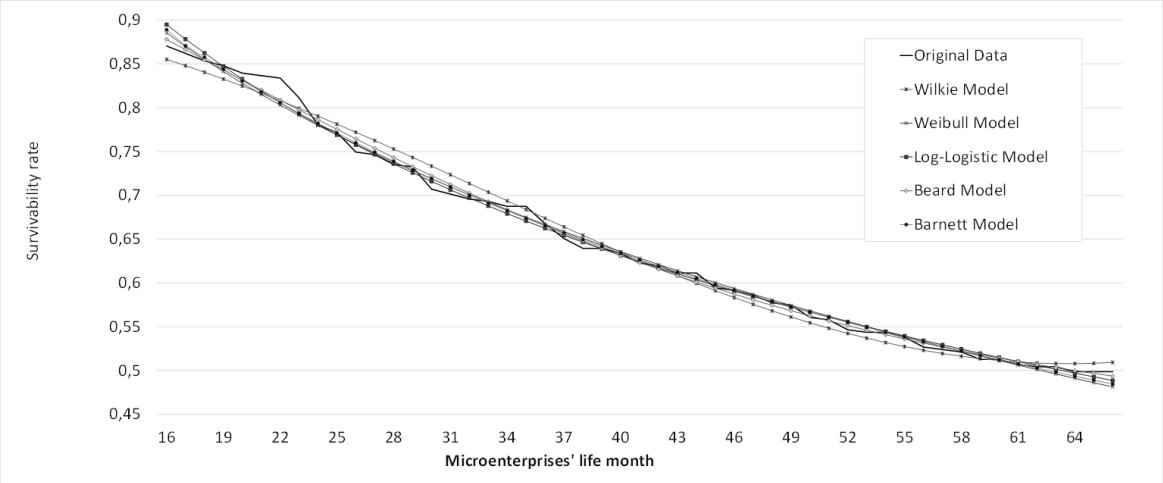


Figure 3 A graphical comparison between original data and the particular model's output (for enterprises lives at least 66 months). Source: own research

References

[1] Aigbavboa, C.O., Tshikhudo, L.M., Thwala, W.D. (2014). Identification of critical success factors for the survival of small, medium and micro enterprise contracting firms in the greater Johannesburg metropolitan area, *Proceedings 8th Construction Industry Development Board*, pp. 349-356, http://hdl.handle.net/10210/12045

[2] Aremu, M. A. (2011). Small and Medium Scale Enterprises as A Survival Strategy for Employment Generation in Nigeria. *Journal of Sustainable Development, Vol. 4 (No. 1),* pp. 200-206, Feb. 2011, https://doi.org/10.5539/jsd.v4n1p200

[3] Balicki, A. (2006). *Analiza przeżycia i tablice wymieralności*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

[4] Baran, J., Kalinowski, H., Leszczyńska, B., Lis, M. (2016). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój I Raport Wskaźnikowy*, Warszawa: Ministerstwo Inwestycji i Rozwoju, https://www.power.gov.pl/media/19541/IBS_I_Raport_Wskaznikowy_07042016.pdf

[5] Beard, R.E. (1971). *Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes, in Biological Aspects of Demography*. W.Brass, ed., Taylor & Francis, London, 1971, http://www.healthsters.com/Beard-1971.pdf

[6] Brüderl, J., Preisendörfer, P., Ziegler, R. (1992). Survival Chances of Newly Founded Business Organizations. *American Sociological Review, Volume 57,* pp. 227-241, https://www.researchgate.net/publication/228271230_Survival_Chances_of_Newly_Founded_Business_Organization

[7] Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics Volume 13(2),* pp. 215–232, https://projecteuclid.org/euclid.aoms/1177731607

[8] Carriere, J. F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries, Vol. 44*, pp. 77-99, https://www.soa.org/library/research/transactions-of-society-of-actuaries/1990-95/1992/january/tsa92v446.pdf

[9] Fertala, N. (2008). The shadow of death: do regional differences matter for firm survival across native and immigrant entrepreneurs?, Empirica. *Journal of European Economics, Volume 35 (1),* pp. 59–80, 2008 https://link.springer.com/article/10.1007/s10663-007-9051-2

[10] Forfar, D. O. (2006). *Mortality laws, Encyclopedia of Actuarial Science,* John Wiley & Sons Ltd 2006, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470012505.tam029

[11] Freeman, J., Carroll, G. R., Hannan, M. T. (1983). The Liability of Newness "Age Dependence in Organizational Death Rates". *American Sociological Review, Volume 48 (5),* pp. 692-710, https://www.jstor.org/stable/pdf/2094928.pdf

[12] Fritsch, M., Brixy, U., Falck, O. (2004). The Effect of Industry, Region and Time on New Business Survival – A Multi-Dimensional Analysis. *Discussion Papers on Entrepreneurship, Growth and Public Policy,*

Max Planck Institute for Research into Economic Systems Group Entrepreneurship, Growth and Public Policy, Jena, ftp://papers.econ.mpg.de/egp/discussionpapers/2004-31.pdf
[13] Gavrilov, L.A., Gavrilova, N.S. (2011). Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal, volume 15 (3),* pp. 432–447, http://longevity-science.org/pdf/Mortality-NAAJ-2011.pdf
[14] Gompertz, B. (1825). 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*, pp. 513-583, https://www.medicine.mcgill.ca/epidemiology/hanley/c609/material/Gompertz-1825.pdf
[15] Gurgul, H., Zając, P. (2016). Zastosowanie modelu przyspieszonej porażki do analizy przeżycia małopolskich przedsiębiorstw. *Przegląd Statystyczny, Volume LXIII (4),* pp. 391-409, http://keii.ue.wroc.pl/przeglad/Rok%202016/Zeszyt%204/2016_63_4_391-409.pdf
[16] Holt, C.C. (1957). *Forecasting trends and seasonals by exponentially weighted averages*, Carnegie Institute of Technology, Pittsburgh 1957, ONR memorandum no. 52, https://doi.org/10.1016/j.ijforecast.2003.09.015 (reprint).
[17] Hyytinena A., Pajarinenb M., Rouvinenb P. (2015). Does innovativeness reduce startup survival rates? *Journal of Business Venturing, Volume 30 (4),* pp. 564–581, http://www.sciencedirect.com/science/article/pii/S0883902614000913
[18] Jørgensen, S.E. (Ed.) Developments in Environmental Modelling. *Fundamentals of Ecological Modelling, vol. 9*, Elsevier, Amsterdam, 1986, https://www.sciencedirect.com/bookseries/developments-in-environmental-modelling/vol/9
[19] Kalleberg, A. L., Leicht, K. T. (1991). Gender and Organizational Performance: Determinants of Small Business Survival and Success. *Academy of Management Journal, Volume 34 (1),* pp. 136-161, https://www.jstor.org/stable/256305
[20] Kalninsa, A., Williams M. (2014). When do female-owned businesses out-survive male-owned businesses? A disaggregated approach by industry and geography. *Journal of Business Venturing, Volume 29 (6),* pp. 822–835, https://www.sciencedirect.com/science/article/pii/S0883902613001092
[21] Lomax, K. S. (1954). Business failures. Another example of the analysis of failure data. *Journal of the American Statistical Association, Vol. 49, No. 268*, pp. 847–52. https://doi.org/10.1080/01621459.1954.10501239
[22] Łobos, K., Szewczyk, M. (2002). Survival analysis: A case study of micro and small enterprises in Dolnośląskie and Opolskie Voivodship (Poland). *Ekonomická revue – Central European Review of Economic Issues, Volume 15*, pp. 207-216, https://www.ekf.vsb.cz/export/sites/ekf/cerei/cs/Papers/VOL15NUM04PAP01.pdf
[23] Makeham, W. (1860). On the Law of Mortality and Construction of Annuity Tables. *The Assurance Magazine and Journal of the Institute of Actuaries, 8(6),* pp. 301-310, https://doi.org/10.1017/S204616580000126X
[24] Richards, S. J. (2010). A handbook of parametric survival models for actuarial use. *Scandinavian Actuarial Journal 1–25(4),* pp. 233-257, https://doi.org/10.1080/03461238.2010.506688
[25] Shane, S. (2012). Small Business Failure Rates by Industry: The Real Numbers. *Small Business Trends, Volume (9),* http://smallbiztrends.com/2012/09/failure-rates-by-sector-the-real-numbers.html
[26] Siler, W. (1979). A Competing-Risk Model for Animal Mortality. *Ecology, vol. 60*, pp. 750-757, https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1936612
[27] Szmit A., Szmit M. (2018). Modeling of survivability of enterprises created under the Human Capital Operational Programme - a case study. *Proceedings of 36th International Conference Mathematical Methods in Economics*, Jindřichův Hradec, pp. 533-538, https://mme2018.fm.vse.cz/wp-content/uploads/2018/09/MME2018-Electronic_proceedings.pdf
[28] Szmit, A., Lisiak-Felicka, D., Szmit, M. (2017). *Badania przeżywalności firm utworzonych w ramach projektów realizowanych z Programu Operacyjnego Kapitał Ludzki. Studium empiryczne dla Fundacji Inkubator*, Wydawnictwo Politechniki Łódzkiej, Łódź 2017, http://repozytorium.p.lodz.pl/bitstream/handle/11652/1484/ASzmit_DLisiak-Felicka_MSzmit_Badania_przezywalnosci_firm.pdf
[29] Tabeau, E. (2001). A Review of Demographic Forecasting Models for Mortality. In: Tabeau E., van den Berg Jeths A., Heathcote C. (Eds.) *Forecasting Mortality in Developed Countries European Studies of Population, vol 9*, Springer, Dordrecht, pp. 1-29, https://doi.org/10.1007/0-306-47562-6_1
[30] van Praag, C. M. (2003). Business Survival and Success of Young Small Business Owners. *Small Business Economics, Vol. 21 (1),* pp. 1-17, https://link.springer.com/article/10.1023%2FA%3A1024453200297

Multiplication Effects of Social and Health Services

Jaroslav Sixta[1], Jakub Fischer[2]

Abstract. The paper deals with expenditures on social and health services in the Czech economy. This study aims at expenditures of Czech households and their impact measured by the input-output analysis. We use the key principles given by Leontief production function and multipliers derived from the static input-output model. We focus mainly on the impact on gross value added, gross domestic product and employment. The structural impacts on the Czech economy are estimated for different scenarios. These scenarios result from changing the share of older people requiring these services. In the current demographic situation, the Czech Republic will be facing ageing, and similar studies should contribute to the debate for the preparation for demographic ageing. The potential of the use of input-output analysis is huge, but it is rarely used for similar purposes in the Czech Republic. However, we also do not hide the limitations of input-output analysis, as well.

Keywords: Input-Output, Ageing, Health, Social, Services.

JEL Classification: C67, J11, J14
AMS Classification: 93D25, 91D20

1 Introduction

Population ageing represented by the increase of people in higher age groups affects all developed countries. Well-known is the situation of Japan and its economic consequences. The combination of the increase in life expectancy and the decrease in the number of births inevitably lead to a challenging situation in many areas of society. In contemporary Europe, so-called demographic ageing is presenting in many countries. This process finally entered the Czech Republic, as well. We put aside the political discussion about optimal retirement age and a fair amount of retirement pensions, and we focus on the current demographic projection only. In our study, we focus on the input-output analysis of ageing effects. We assume that the problem of ageing does not connect with retirement pensions. It very closely connects with the pressure on health and social services. In other words, we will need more doctors, nurses, social workers etc.

In our paper, we prepare easy economic prediction by 2060 for the Czech Republic. We combine simple static input-output analysis and official demographic projections published by the Czech Statistical Office [1]. Concerning the complexity of the issue, we can recommend Computable General Equilibrium (CGE) models, see [5]. However, for our purposes, a simple input-output model fits well. We can estimate induced change of output and employment due to the increased demand for medicines, health and social services.

There is a lack of Czech literature devoted to the models of ageing. Some works were devoted to the availability of services, see [4] and some to the estimates of monetary effects such as [3]. On the international level, the situation is much better since many scientists are dealing with the ageing issue, see [8] or [6]. Skirbekk [10] also mentions the question of individual productivity. Richter [9] and Okon [7] also elaborate on the issue of the readiness of the population for ageing where the necessary health care staff in Sachsen is estimated. Our approach respects similar procedures as in literature since we focus on macroeconomic impacts only. We put aside effects such as social burden or individuals' preferences.

2 Data and Methodology

We use standard tool for modelling economic impacts based on the structure of the current economy, input-output analysis in its static form. Thus we do not estimate dynamic multiplication effects. The reason is that we want to illustrate the changes between the current state of the art and possible future development. Our approach is in line with standard usage of these relatively simple econometric models. One has been aware that Leontief production function fixes the possibility of substitution of inputs. In other words, economic progress may bring significant improvements in medicine or social services that the rate of human labour will be decreasing.

[1] University of Economics, Department of Economic Statistics, W. Churchilla 4, Prague, sixta@vse.cz
[2] University of Economics, Department of Economic Statistics, W. Churchilla 4, Prague, fischerj@vse.cz

On the contrary, we can expect that industries of medicine production, health care and social care cannot avoid human labour. Despite the primary input (labour or capital), the output of these industries will be rising. And this is the aim we would like to illustrate.

The data comes from the Czech Statistical Office. We combine national accounts' data, mainly symmetric-input-output tables (lates available are for 2015 at current prices) and data from demography. We also use demographic projections (see [1]) compiled in three variants. These variants comprise low, middle and high scenario depending on the main demographic parameters.

We use the model based on standard input-output analysis and symmetric input-output tables for the use of domestic products:

$$\Delta \mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1} \Delta \mathbf{y} \tag{1}$$

where

x.. vector of output,

A.. Leontief matrix of technical coefficients,

y.. final demand.

The economic impact measured is expressed in **y** as final demand change. We assume that current expenditures for medicine, health and social services are proportionally fixed to the structure of the population. In other words, we estimate an increase in the demand for services about a different scenario. We can hardly expect that the rise of the share of the older population will require such services in the same proportion. On the contrary, we need some variant based inputs. The health condition of people is continuously improving, technological processes play a crucial role in treatments and preventions, etc. According to health insurance companies, their expenditures takes about 38% for people over 65 years in 2016, see CZSO [2].

3 Economic and Demographic Background

Even though the share of older people has been continuously rising in the Czech Republic, the significant increase is expected in the next 20 years. We do not tackle the issue of fair or reasonable retirement pension rate of statutory retirement pension age; we focus on the services necessary to be provided. Currently, we have less than 20% of people at the age of 65 and higher. Around the year 2040, we expect a dramatical increase on this share depending on the variant. In one of the worst scenario (low variant), this share should exceed 30% of the total population, see Figure 1.

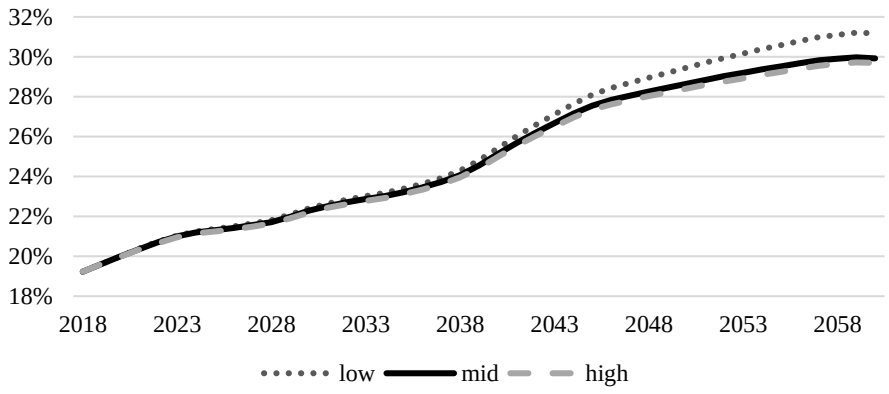


Figure 1 The share of people over 65 years

It is a serious question if the Czech society is ready for such an increase in older people. On issue is the retirement pension and if the state or private funds can ensure the life in dignity. However, we think, that serious issue is also the availability of health and social services and their funding, as well. This issue is the keystone of ageing. Lots of people will be able to work longer, but more people will require some permanent or temporary assistance even they are alive. The concept of healthy life expectancy is very serious, and we cannot omit it but we focus on the macroeconomic impact only.

4 Input-Output Analysis of Ageing Driven Consumption

We assume that the change in the structure of the population will cause the changes in output and employment necessary to satisfy different consumers' needs. We estimate the overall impact by statistic input-output analysis where we derive the change of the final use from the number of the consumers (users). The number of users of medicines and health and social services will be rising as the population is ageing. We provide estimates in three variants in line with the available demographic projection, see Table 1. Low variant leads to the increase of final demand of specific commodities by 23.8% and an overall change of the output by 0.3%. When using middle or high variants, the increase of final demand is 26.1%, 28.0% respectively. The increase in output is about 0.4%, 0.4% respectively. In nominal terms, the overall output will be change between 36 and 42 billion CZK at prices of 2015.

The impact of different consumer structures on macroeconomic figures in 2060 is more serious. Depending on the variant, the increase of final demand can reach 53.1%, 67.7% or 79.2%. Similarly, the change in the output is ranging from 0.7% to 1.1%. In nominal terms, it means that the final demand will change by 58, 74 or 87 CZK billion. The subsequent increase in output reaches 79, 101, 118 CZK billion. The difference between high and low variant is rising in comparison with the situation in 2030. That is due to the different development of the share of people over 65 since we estimate 471 thousand people in a low variant and 552 thousand people in a high variant.

	2030			2060		
	Final demand	**Total output**	**Employment**	**Final demand**	**Total output**	**Employment**
Low	23.8%	0.3%	9.1%	53.1%	0.7%	20.4%
Medium	26.1%	0.4%	10.0%	67.7%	0.9%	26.0%
High	28.0%	0.4%	10.7%	79.2%	1.1%	30.4%

Table 1 Changes due to consumption

The similar situation stands for employment in the industry of health and social services. It the increase in necessary staff is very significant. In 2030, the increase can reach from 25.6% to 30.1% and from 57% to 85.1% in 2060. These impacts are much more serious than impacts measured in monetary values. Preparation of qualified staff ranging from doctors, nurses and other works is a long-term process.

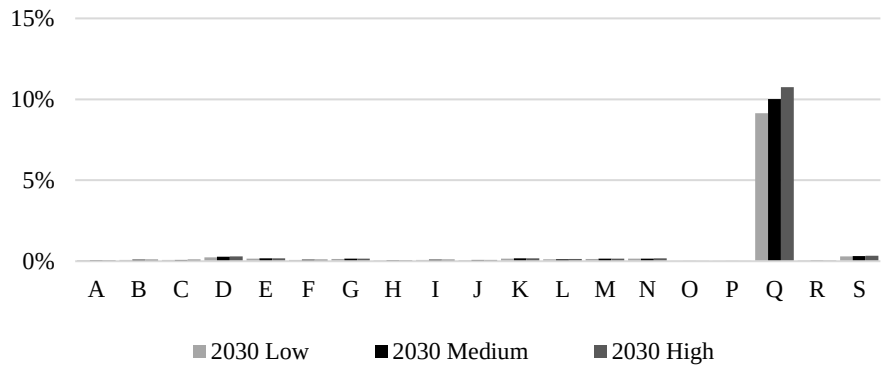


Figure 2 Relative changes in output in 2030

The overall impacts on output can be split into the level of industries, see Figure 2 and Figure 3. The impact of consumer preferences is distributed across all industries, but the most affected industries are health and social care (Q)[3]. In 2030, the increase in the output of these industries may range between 9% and 11% and 20% and 30% in 2060.

[3] Codes of the classification of industries are following: : A - Agriculture, forestry and fishing, B - Mining and quarrying, C – Manufacturing, D - Electricity, gas, steam and air conditioning supply, E - Water supply; sewerage, waste management and remediation activities, F – Construction, G - Wholesale and retail trade; repair of motor vehicles and motorcycles, H - Transportation and storage, Other: I - Accommodation and food service activities, J - Information and communication, K - Financial and insurance activities, L - Real estate activities, M - Professional, scientific and technical activities, N - Administrative and support service activities, O - Public administration and defence; compulsory social security, P – Education, Q - Human health and social work activities, R - Arts, entertainment and recreation, S - Other service activities, T - Activities of households as employers and producers for own use

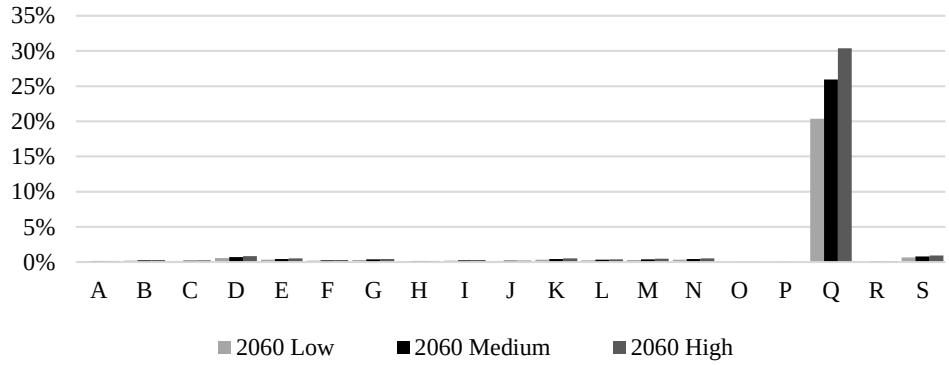


Figure 3 Relative changes in output in 2060

5 Conclusion

The impact of ageing on macroeconomic figures such as output and employment is undisputable. The real question is only how big the influence will be. How society and its economy will be changed. Our approach offers easy, and brief analysis of the changes induced by consumers' preferences drive ageing. We assume that an increased share of the population over 65years will demand different mix products. In this way, it means that people will live longer, but they will need more health and social treatment during their long life. Some people will require full-time permanent medical treatment or attention and some of them part-time or temporary, which requires deep economic analysis. In our example, we rely on simple input-output analysis, and we do not consider any social or economic progress. Thus, we consider our models and results as a limiting worst scenario. Despite that, the results are alarming.

In comparison with 2015, we estimate that an increased number of people over 65 may lead to a 10% increase in employment in 2030 in industries of health and social care. In 2060, this increase can reach 26%. These figures represent several doctors, nurses and social workers. It is obvious that increased final demand (consumption expenditures) leads to the changes in output and employment. We present our estimates at prices of 2015 since we use symmetric input-output tables for 2015. We use three variants of official demographic projections and the differences between them are not very significant in 2030.

On the contrary, the differences in 2060 are very significant. We also present the impacts on total output and output by industries. However, the most serious issue is the impact on employment since we can hardly expect the health and social care can be fully automatized. We expect that human labour will be an important production factor in these industries in the future, as well. The question is just the share. Threatening is the missing national strategy for overcoming these effects.

Acknowledgements

Supported by the Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague and the project "Economy of Successful Ageing", no. 19-03984S.

References

[1] Czech Statistical Office (2018a). *Projekce obyvatelstva České republiky* 2018 - 2100. Praha.
[2] Czech Statistical Office (2018b). *Focus on Women and Men*. Praha.
[3] Janíčko M. & Tsharakyan A. (2013). K udržitelnosti průběžného důchodového systému v kontextu stárnutí populace v České republice. *Politická ekonomie,* vol. 61 no. 3 (pp. 321-337). Praha
[4] Langhamrová, J., Šimková, M. & Sixta, J. (2018). Makroekonomické dopady rozšiřování sociálních služeb pro stárnoucí populaci České republiky. *Politická ekonomie*, vol. 66 no. 2 (pp. 240-259). Praha
[5] Lisenkova, K., McGregor, P.G., Pappas, N., Swales, J.K., Turner, K. & Wright, R. (2010). Scotland the Grey: A Linked Demographic-Computable General Equilibrium (CGE) Analysis of the Impact of Population Ageing and Decline, *Regional Studies*, 44 (pp. 1351-1368).

[6] Mahlberg, B., Freund, I., Cuaresma, J. C. & Prskawetz, A. (2013). Ageing, productivity and wages. *Labour Economics*, 22, (pp. 5-15). Austria.
[7] Okon, J. (2015). Schätzung der Beschäftigungspotenziale im sächsischen Gesundheitswesen bis 2030, S*tatistisches Landesamt des Freistaates Sachsen*, no. 2. Sachsen
[8] Prskawetz, A., Fent, T., Barthel, W., Crespo-Cuaresma, J., Lindh, T., Malmberg, B. & Halvarsson, M. (2007). The relationship between demographic change and economic growth in the EU. *Research Report No. 32, Vienna Institute for Demography*, Austrian Academy of Science. Austria.
[9] Richter, B. (2018). Aktualisierte Berechnung des voraussichtlichen Bedarfs, *Sächsischen Gesundheitswesen 2030*. Satistisches Landesamt des Freistaates Sachsen. Sachsen.
[10] Skirbekk, V. (2008). Age and productivity potential: A new approach based on ability levels and industry-wide task demand. *Population and Development Review*, 34, (pp. 191-207).

Second order stochastic dominance constraints in decision dependent randomness portfolio optimization problems

Miloš Kopa[1]

Abstract. The paper deals with stochastic portfolio optimization problems which maximize a given functional under second - order stochastic dominance constraints in presence of endogenous randomness. Endogenous randomness (or decision dependent randomness) means that the probability distribution of asset returns may depend on the decision variables, i.e. on the weights associated to the assets. This may occur typically in the high frequency trading or in the illiquid markets, when a massive investment of one investor may attract others investors, at least for a small time period. Firstly, we modify the classical second-order stochastic dominance relation between returns of two given portfolios for the case with endogenous randomness of returns. Secondly, we apply this new constraint to the portfolio optimization problem. Finally, we present an example demonstrating the differences in optimal portfolios when endogenous randomness is omitted (exogenous randomness is assumed).

Keywords: Decision dependent randomness, second order stochastic dominance, portfolio optimization

JEL Classification: D81, G11
AMS Classification: 91B16, 91B30

1 Introduction

Randomness in the stochastic portfolio selection models may be considered in two ways: endogenous and exogenous. Endogenous randomness is inner uncertainty of the model. It means that the random (uncertain) element of the problem may depend on the solution (decision). For example, the decision-maker can force one realization to become more probable by his/her decision, or the whole distribution can be shifted as the consequence of his/her decision. Exogenous uncertainty is entering the model from outside. In this case, the decision-maker cannot affect the underlying probability distribution. The previous work on the endogenous uncertainty in optimization problems is limited to a few papers only. Their reviews can be found e.g. in [29], [5] or [9].

The most commonly used stochastic dominance relations are so called the first-order and the second-order stochastic dominance. This is due to clear and easy interpretation of these rules. They allow comparing random variables, in financial applications random returns or losses of assets (portfolios). In general, we say that a random variable X dominates a random variable Y with respect to the N-th order stochastic dominance, $N = 1, 2$ if

$$\mathbb{E}u(X) \geq \mathbb{E}u(Y) \text{ for all } u \in U_N$$

where the set of admissible utility functions for the N-th order stochastic dominance is defined as follows:

$$U_N = \{u(x) \in D_N : (-1)^k u^{(k)} \leq 0, \ \forall x, \ k = 1, ..., N\}$$

where D_N is the class of N-times differentiable functions and $u^{(k)}$ stands for the k-th derivative of function u. The basics of stochastic dominance go back to 1960th, see [24], [10], [11] or [27]. See [17] for more details. In the last decade, a substantial development of the stochastic dominance applications in finance was observed. In particular:

- necessary and sufficient conditions for portfolio efficiency with respect to stochastic dominance criteria were derived and applied, see e.g. [22], [8], [15], [21]
- portfolio enhancement using stochastic dominance rules proved to be a promising way comparing to classical mean-risk models, see e.g. [26], [12], [23], [14], [16], [28].
- Special Data Envelopment Analysis models wich are equivalent to portfolio efficiency tests with respect to stochastic dominance were presented in [1], [2], [3].

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

- more robust version of stochastic dominance relations and stochastic dominance efficiency were proposed, see e.g. [4], [13], [6] or [7].
- more general stochastic dominance (ordering) rules were introduced and analyzed, see e.g. [19], [18] and references therein, or [20].

Since the most popular stochastic dominance relation is the second-order stochastic dominance, in this paper we focus on the portfolio selection problems which maximize expected return under the second-order stochastic dominance constraints. First, we formulate a portfolio optimization model with the second-order stochastic dominance constraints under assumption of exogenous randomness. Then we modify the model for the case of endogenous randomness. For both cases we present some tractable formulations under assumption of particular distributions of returns. In order to present the ideas in the simplest way, we focus on the Gaussian distributions and discrete distributions where the realizations of the random variables are equiprobable. Finally, we present a numerical example showing how the optimal portfolios may differ if endogenous randomness is incorporated in the problem.

The remainder of this paper is structured as follows. Section 2 presents a notation and basic properties of the SSD relation. It is followed by a summary of portfolio selection models (with exogenous or endogenous randomness) applying SSD constraints in Section 3. These models assume an empirical or Gaussian distribution of the returns. Section 4 shows a numerical example. The paper is concluded in Section 5.

2 Preliminaries and notations

We consider a random vector $\mathbf{r} = (r_1, r_2, ..., r_N)'$ of returns of N assets. We will use $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)'$ for a vector of portfolio weights and the portfolio possibilities are given by

$$\Lambda = \{\boldsymbol{\lambda} \in R^N | \mathbf{1}'\boldsymbol{\lambda} = 1, \ \lambda_n \geq 0, \ n = 1, 2, \ldots, N\}.$$

Let $F_{\mathbf{r}'\boldsymbol{\lambda}}(x)$ denote the cumulative probability distribution function of returns of portfolio $\boldsymbol{\lambda}$ and

$$F^{[2]}_{\mathbf{r}'\boldsymbol{\lambda}}(t) = \int_{-\infty}^{t} F_{\mathbf{r}'\boldsymbol{\lambda}}(x)dx.$$

The same notation applies for portfolio τ.

Definition 1. Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the second order stochastic dominance ($\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau}$) if

$$F^{[2]}_{\mathbf{r}'\boldsymbol{\lambda}}(t) \leq F^{[2]}_{\mathbf{r}'\boldsymbol{\tau}}(t) \qquad \forall t \in \mathbb{R}.$$

This relation is sometimes called a weak SSD and the equivalent definition, presented in e.g. [11], [17] or [15] is based on comparison of expected utility of portfolio returns:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow \mathbb{E}u(\mathbf{r}'\boldsymbol{\lambda}) \geq \mathbb{E}u(\mathbf{r}'\boldsymbol{\tau})$$

for all utility functions $u \in U_2$.

If we allow that $\mathbf{r}$ may depend on the decision vector $\boldsymbol{\lambda}$ we have to make Definition 1 more general.

Definition 2. Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the second order stochastic dominance with endogenous randomness ($\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau}$) if

$$F^{[2]}_{\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda}}(t) \leq F^{[2]}_{\mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau}}(t) \qquad \forall t \in \mathbb{R}.$$

2.1 Gaussian distribution of returns

Assume that random vector of returns has a multivariate Gaussian distribution: $\mathbf{r} \sim N(\boldsymbol{\mu}, \Sigma)$. Then, following [17]:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow (\boldsymbol{\mu}'\boldsymbol{\lambda} \geq \boldsymbol{\mu}'\boldsymbol{\tau} \wedge \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda} \leq \boldsymbol{\tau}'\Sigma\boldsymbol{\tau})$$

One of the easiest decision dependence is when only the parameters of the distribution depend on the solution but not its type. In this case, we assume that random vector of returns still has a multivariate Gaussian distribution, however now with different parameters: $\mathbf{r}(\boldsymbol{\lambda}) \sim N(\boldsymbol{\mu}(\boldsymbol{\lambda}), \Sigma(\boldsymbol{\lambda}))$. Then portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the second order stochastic dominance with endogenous randomness if and only if:

$$\boldsymbol{\mu}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \geq \boldsymbol{\mu}(\boldsymbol{\tau})'\boldsymbol{\tau} \wedge \boldsymbol{\lambda}'\Sigma(\boldsymbol{\lambda})\boldsymbol{\lambda} \leq \boldsymbol{\tau}'\Sigma(\boldsymbol{\tau})\boldsymbol{\tau}.$$

2.2 Empirical distribution of returns

Assume now that the returns of the assets for the various scenarios (taken with the same probability) 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.

Following [6] we may formulate the SSD constraints using Conditional Value at Risks as follows:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow \mathrm{CVaR}_\alpha(-\mathbf{r}'\boldsymbol{\lambda}) \geq \mathrm{CVaR}_\alpha(-\mathbf{r}'\boldsymbol{\tau})$$

for all $\alpha = 0, 1/T, ..., (T-1)/T$, where $\mathrm{CVaR}_\alpha(-\mathbf{r}'\boldsymbol{\lambda})$ is given by optimization formula presented in [25].

Alternatively, the necessary and sufficient condition for SSD can be rewritten using a double stochastic matrix P as follows: $\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow$

$$\exists P = \{p\}_{i,j=1}^T : X\boldsymbol{\lambda} \geq PX\boldsymbol{\tau}, \mathbf{1}'P = \mathbf{1}, P\mathbf{1} = \mathbf{1}, p_{i,j} \geq 0, i, j = 1, 2, ..., T. \tag{1}$$

If endogenous randomness affects only the values of the return scenario matrix X but not the probabilities of the scenarios we may modify (1)-(2) as follows: $\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau} \Longleftrightarrow$

$$\exists P = \{p\}_{i,j=1}^T : X(\boldsymbol{\lambda})\boldsymbol{\lambda} \geq PX(\boldsymbol{\tau})\boldsymbol{\tau}, \mathbf{1}'P = \mathbf{1}, P\mathbf{1} = \mathbf{1}, p_{i,j} \geq 0, i, j = 1, 2, ..., T. \tag{2}$$

where $X(\boldsymbol{\lambda})$ is the scenario return matrix if portfolio $\boldsymbol{\lambda}$ is chosen. Similarly, if we invest in portfolio $\boldsymbol{\tau}$ the scenario return matrix is $X(\boldsymbol{\tau})$.

3 Portfolio optimization models

In this paper, we shall deal with with portfolio selection models maximizing mean return under second order stochastic dominance constraints. If exogenous randomness is assumed, the model takes the form:

$$\begin{aligned} & \max \mathbb{E}(\mathbf{r}'\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \\ & \boldsymbol{\lambda} \in \Lambda \end{aligned} \tag{3}$$

where $\boldsymbol{\tau}$ is a given benchmark (reference) portfolio.

3.1 Gaussian distribution of returns

If we consider a Gaussian distribution of returns, following section 2.1, the model (3) simplifies to:

$$\begin{aligned} & \max \boldsymbol{\mu}'\boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\mu}'\boldsymbol{\lambda} \geq \boldsymbol{\mu}'\boldsymbol{\tau} \\ & \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda} \leq \boldsymbol{\tau}'\Sigma\boldsymbol{\tau} \\ & \boldsymbol{\lambda} \in \Lambda \end{aligned} \tag{4}$$

and the endogenous randomness modification takes the form:

$$\begin{aligned} & \max \boldsymbol{\mu}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\mu}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \geq \boldsymbol{\mu}(\boldsymbol{\tau})'\boldsymbol{\tau} \\ & \boldsymbol{\lambda}'\Sigma(\boldsymbol{\lambda})\boldsymbol{\lambda} \leq \boldsymbol{\tau}'\Sigma(\boldsymbol{\tau})\boldsymbol{\tau} \\ & \boldsymbol{\lambda} \in \Lambda. \end{aligned} \tag{5}$$

since only the parameters of the probability distribution depend on the decision.

3.2 Empirical distribution of returns

If we consider a discrete distribution of returns with equiprobable scenarios (atoms) we can follow (1) and formulate (3) as a linear programming problem:

$$\begin{aligned} & \max \mathbb{E}(\mathbf{r}'\boldsymbol{\lambda}) \\ \text{s.t.} \quad & X\boldsymbol{\lambda} \geq PX\boldsymbol{\tau} \\ & \mathbf{1}'P = \mathbf{1} \\ & P\mathbf{1} = \mathbf{1} \\ & p_{i,j} \geq 0, i, j = 1, 2, ..., T \\ & \boldsymbol{\lambda} \in \Lambda \end{aligned} \tag{6}$$

In the case, that endogenous randomness affects only the values of the return scenario matrix X but not the probabilities of the scenarios, following (2) we get:

$$\begin{aligned} & \max \mathbb{E}(\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda}) \\ \text{s.t.} \quad & X(\boldsymbol{\lambda})\boldsymbol{\lambda} \geq PX(\boldsymbol{\tau})\boldsymbol{\tau} \\ & \mathbf{1}'P = \mathbf{1} \\ & P\mathbf{1} = \mathbf{1} \\ & p_{i,j} \geq 0, i, j = 1, 2, ..., T \\ & \boldsymbol{\lambda} \in \Lambda \end{aligned} \tag{7}$$

4 Numerical example

In this section we assume only three assets: $N = 3$ and the benchmark portfolio is $\boldsymbol{\tau} = (0, 0, 1)$.

4.1 Gaussian distribution of returns

Let's start with the case of exogenous randomness when the parameters of the Gaussian distribution do not depend on the selected portfolio $\boldsymbol{\lambda}$: assume that $\boldsymbol{\mu} = (2\ \ 3\ \ 4)'$ and

$$\Sigma = \begin{pmatrix} 6 & -3 & 0 \\ -3 & 12 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

Since the benchmark portfolio has mean return equal to 4, which is the highest possible, the optimal solution of (4) is $\boldsymbol{\lambda}^* = \boldsymbol{\tau}$.

Now assume that if a massive investment is done in the first asset ($\lambda_1 \geq 0.9$) then: $\mu_1 = 2+30(\lambda_1-0.9)$; if in the second asset ($\lambda_2 \geq 0.9$) then: $\mu_2 = 3+20(\lambda_2-0.9)$ and if in the third asset ($\lambda_3 \geq 0.9$) then: $\mu_3 = 4+5(\lambda_3-0.9)$. Moreover, assume that the variance-covariance matrix Σ is not affected by the decision. In this case, the mean return of the benchmark portfolio equals to 4.5 and the optimal solution of (5) is $\boldsymbol{\lambda}^* = (1, 0, 0)$.

4.2 Empirical distribution of returns

Assume that the empirical distribution has only 3 scenarios and the scenario matrix (in percentage) is:

$$X = \begin{pmatrix} 3 & 8 & 5 \\ 2 & 4 & 3 \\ 11 & 3 & 6 \end{pmatrix}$$

Let's solve first the problem (6). Since the first asset is the most profitable, we would like to invest in it as much as possible. Since the minimal return of the benchmark portfolio is 3% the maximal possible (and optimal) weight of the first asset is 0.5 when combined with the second asset. Hence, the optimal portfolio is $\boldsymbol{\lambda}^* = (0.5, 0.5, 0)$.

Now assume that if a massive investment is done in the first (the second, the third) asset, that is $\lambda_1 \geq 0.9$ ($\lambda_2 \geq 0.9$, $\lambda_3 \geq 0.9$) the return of the first (the second, the third) asset is increased by 1 for the last scenario. Consider the following subsets of Λ:

$$\Lambda_1 = \{\boldsymbol{\lambda} \in \Lambda : \lambda_1 \geq 0.9\}, \quad \Lambda_2 = \{\boldsymbol{\lambda} \in \Lambda : \lambda_2 \geq 0.9\}, \quad \Lambda_3 = \{\boldsymbol{\lambda} \in \Lambda : \lambda_3 \geq 0.9\}.$$

Now we can express the decision dependent random returns:

$$X(\boldsymbol{\lambda}) = \begin{pmatrix} 3 & 8 & 5 \\ 2 & 4 & 3 \\ 11 & 3 & 6 \end{pmatrix} + \mathcal{I}(\boldsymbol{\lambda} \in \Lambda_1) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \mathcal{I}(\boldsymbol{\lambda} \in \Lambda_2) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \mathcal{I}(\boldsymbol{\lambda} \in \Lambda_3) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In this case, problem (7) can be decomposed in four separate problems which are very similar to each other, just the last constraint of (7) is modified as follows:

1. $\boldsymbol{\lambda} \in \Lambda_1$
2. $\boldsymbol{\lambda} \in \Lambda_2$
3. $\boldsymbol{\lambda} \in \Lambda_3$
4. $\boldsymbol{\lambda} \in \Lambda \backslash (\Lambda_1 \cup \Lambda_2 \cup \Lambda_3)$.

When the last constraint of (7) is replaced by $\boldsymbol{\lambda} \in \Lambda_1$, no feasible solution exists, because the smallest return of any portfolio $\boldsymbol{\lambda} \in \Lambda_1$ is smaller than the smallest return of the benchmark. In the second case (when $\boldsymbol{\lambda} \in \Lambda_2$), the optimal portfolios are $\boldsymbol{\lambda}^2 = (1-k, k, 0)$, $k \in [0.9, 1]$ and the optimal objective value is $16/3$. The third modification (when $\boldsymbol{\lambda} \in \Lambda_3$) finds the optimal solution $\boldsymbol{\lambda}^3 = (0.05, 0.05, 0.9)$ with optimal objective value smaller than $16/3$. Finally, the last possibility ($\boldsymbol{\lambda} \in \Lambda \backslash (\Lambda_1 \cup \Lambda_2 \cup \Lambda_3)$) gives the same optimal solution as the exogenous randomness case $\boldsymbol{\lambda}^4 = (0.5, 0.5, 0)$ and its optimal objective value is $31/6$. Since $16/3 > 31/6$ the optimal solutions of the endogenous randomness portfolio selection model are $\boldsymbol{\lambda}^* = (1-k, k, 0)$, $k \in [0.9, 1]$.

5 Conclusions

In this paper we analyzed the portfolio selection models with the second-order stochastic dominance constraints under assumption of the endogenous randomness. It means that the multivariate distribution of random returns depends on the decision vector (portfolio weights). We first introduced a definition of SSD between two portfolios under assumption of endogenous randomness and then we formulated the portfolio selection models assuming (i) Gaussian distribution with parameters depending on the decision or (ii) empirical distribution when only the values of scenarios may depend on the decision but not the probabilities of the scenarios - the scenarios are assumed to be equiprobable. Finally, we illustrated a difference between endogenous and exogenous randomness portfolio selection problems on toy examples. Despite the fact that in both problems the endogenous randomness is expressed in a very simple way, the optimal portfolios of the endogenous randomness problems very much differ from the ones of the exogenous randomness problems. Therefore, omitting the endogenous feature of the returns would lead to completely different conclusions.

For the future research, this study can be improved in various ways. For example, one can consider more robust stochastic dominance relations following [13], [6] or [7], using contamination techniques and the worst-case approach. Alternatively, one can compare these results also with the case when the first or the third order stochastic dominance is used.

Acknowledgements

The paper was supported by the grant No. 18-05631S of the Czech Science Foundation.

References

[1] Branda, M., and Kopa, M.: DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* **62** (2012), 106–124.

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

[3] Branda, M., and Kopa, M.: DEA models equivalent to general N-th order stochastic dominance efficiency tests. *Operations Research Letters* **44** (2016), 285–289.

[4] Dentcheva, D., and Ruszczyński, A.: Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming, Series B* **123** (2010), 85–100.

[5] Dupačová, J.: Optimization under exogenous and endogenous uncertainty. In: Proc. of MME06 (L. Lukáš, ed.), University of West Bohemia 2006, 131-136.

[6] Dupačová, J., and Kopa, M.: Robustness in stochastic programs with risk constraints. *Annals of Operations Research* **200** (2012), 55–74.

[7] Dupačová, J., and Kopa, M.: Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research* **234** (2014), 434–441.

[8] Grechuk, B.: A simple SSD-efficiency test. *Optimization Letters* **8** (2014), 2135–2143.

[9] Grossmann, I.E., Apap, R.M., Calfa, B.A., García-Herreros, P. and Zhang, Q.: Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. *Computers and Chemical Engineering* **91** (2016), 3–14.

[10] Hadar, J., and Russell, W. R.: Rules for Ordering Uncertain Prospects. *American Economic Review* **59** (1969), 25–34.

[11] Hanoch, G., and Levy, H.: The efficient analysis of choices involving risk. *Review of Economic Studies* **36** (1969), 335–346.

[12] Hodder, E., Jackwerth, J.C., and Kolokolova, O.: Improved Portfolio Choice Using Second-Order Stochastic Dominance. *Review of Finance* **19** (2015), 1623–1647.

[13] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency. *Kybernetika* **46** (2010), 488–500.

[14] Kopa, M., Moriggia, V. and Vitali, S.: Individual optimal pension allocation under stochastic dominance constraints. *Annals of Operations Research* **260** (2018), 255–291.

[15] Kopa, M., and Post, T.: A general test for SSD portfolio efficiency. *OR Spectrum* **37** (2015), 703–734.

[16] Moriggia, V., Kopa, M. and Vitali, S.: Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega* **87** (2019), 127–141.

[17] Levy, H.: *Stochastic dominance: Investment decision making under uncertainty*. Second edition, Springer, New York, 2006.

[18] Ortobelli, S., Lando, T., Petronio, F., and Tichy, T.: Asymptotic stochastic dominance rules for sums of i.i.d. random variables. *Journal of computational and applied mathematics* **300** (2016), 432–448.

[19] Ortobelli, S., Shalit, H., and Fabozzi, F.: Portfolio selection problems consistent with given preference orderings. *International Journal of Theoretical and Applied Finance* **16** (2013), 38p.

[20] Post, T.: Standard Stochastic Dominance. *European Journal of Operational Research* **248** (2016), 1009–1020.

[21] Post, T., Fang, Y., and Kopa, M.: Linear Tests for DARA Stochastic Dominance. *Management Science* **61** (2015), 1615–1629.

[22] Post, T., and Kopa, M.: General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research* **230** (2013), 321–332.

[23] Post, T., and Kopa, M.: Portfolio Choice based on Third-degree Stochastic Dominance. *Management Science* **63** (2017), 3381–3392 .

[24] Quirk, J. P., and Saposnik, R.: Admissibility and measurable utility functions. *Review of Economic Studies* **29** (1962), 140–146.

[25] Rockafellar, R.T. and Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* **26** (2002), 1443–1471.

[26] Roman, D., Mitra, G., and Zverovich, V.: Enhanced indexation based on second-order stochastic dominance. *European Journal of Operational Research* **228** (2013), 273–281.

[27] Rothschild, M., and Stiglitz, J.E.: Increasing risk: I. A definition. *Journal of Economic Theory* **2** (1970), 225–243.

[28] Rusý, T. and Kopa, M.: An asset liability management stochastic program of a leasing company. *Kybernetika* **54** (2018), 1247–1263.

[29] Sahinidis, N. V.: Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering* **28** (2004), 971–983.

School Performance and Various Socioeconomic Factors: A GWR Approach for Slovak Data

Michaela Chocholatá[1]

Abstract. To ensure the high quality of education process, to enhance the level of knowledge and skills as well as to eliminate the regional disparities in pupil's learning effectiveness and in school performance belong to the main objectives of the national development strategy in Slovakia. This paper deals with the regional performance of primary schools in Slovakia assessed by the external testing of pupils of the 5th year of primary schools on knowledge of mathematics and national language. The regional school performance is further analysed based on the average percentage Maths scores achieved in 2018. Besides revealing the regional spatial patterns in school performance the main aim of this paper is oriented on both the global and local regression analysis assessing the impact of selected socioeconomic factors (average nominal monthly wage, unemployment rate and index of economic dependence of young people) onto the average percentage Maths scores across the Slovak districts. The classic global regression analysis is followed by local regression analysis based on the geographically weighted regression approach which enables to catch the spatial variations in the modelled relationship.

Keywords: school performance, socioeconomic factors, geographically weighted regression

JEL Classification: I25, C21
AMS Classification: 91B72, 62P20

1 Introduction

The issue of education on different levels plays in general one of the key roles in national development strategies. Analysing of pupil's learning effectiveness and of school performance is an interesting issue for parents, for the government as well as for researchers. It is commonly known that the school performance is different from school to school, there are schools performing well and schools performing not so well. Since parents try to find the best school for their children, the effort of government is oriented on improvement of learning programs and strategies to enhance the level of knowledge and skills and to eliminate the inequalities. Researchers are looking for causes of differences in quality of school performance and learning effectiveness.

There have been published many studies analysing the school performance studying the impact of various pupil-level variables and school-level variables as well as of socioeconomic variables (see e.g. [5], [6], [8], [11], [13]). However, as pointed out by Fotheringham, Charlton and Brunsdon [8], Qiu and Wu [13] and Christie [11], it is important not only to analyse the relationships between school performance and various factors, but also to incorporate the association of the analysed variable with the geography to assess the spatial variations in these relationships.

In Slovakia, the movement of the education towards a higher quality is supported by the external testing of the primary and secondary school pupils (Testing 5, Testing 9 and external part of secondary school graduation exam), which provides the government with a very realistic picture of the strengths and weaknesses of both pupils and the learning process. The results of testing can also deal as input data in order to assess the added value in education process, which is an important factor in measuring of school quality [12].

The aim of this paper is to provide both the global and local regression analysis assessing the impact of selected socioeconomic factors (average nominal monthly wage, unemployment rate and index of economic dependence of young people) onto the Testing 5-2018 (T5-2018) Maths test scores[2] across the Slovak districts. Since the global regression analysis does not enable to catch the spatial variations in the modelled relationship, the local regression analysis based on the geographically weighted regression (GWR) approach is used as well.

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, michaela.chocholata@euba.sk.

[2] T5 is based on testing of pupils' knowledge of mathematics and national language. In accordance with Fotheringham, Charlton and Brunsdon [8] we decided to analyse those of Maths.

The paper is organised as follows: section 2 deals with the methodology, section 3 contains data description and the empirical results and section 4 concludes with suggestions for future research.

2 Methodology

To assess the global relationship between the school performance and the socioeconomic variables a classic linear regression model using the ordinary least squares (OLS) method is usually being estimated. However, as pointed out by many researchers (see e.g. [13]), the application of OLS on spatial data usually leads to the violation of statistical assumption of independent residuals and often also to the violation of residual constant variance assumption. In order to analyse the spatial data, it is thus important to use spatial econometric models and methods regarding the spatial dependence and/or spatial heterogeneity of analysed regions. While the spatial autocorrelation indicates that there is a significant tendency towards clustering of similar (dissimilar) values in space, the spatial heterogeneity indicates that the relationship between variables is not consistent across the whole analysed area, i.e. that parameters can vary across group of regions or in a more general case even across individual regions ([3], [9], [13]). As pointed out e.g. by Anselin [1], the above-mentioned spatial effects (spatial autocorrelation and spatial heterogeneity) often appear together and it is not easy to determine whether the spatial effects are in the form of spatial autocorrelation or spatial heterogeneity. The GWR approach developed by Brunsdon, Charlton and Fotheringham (see e.g., Fotheringham, Brunsdon and Charlton [7]) can be used to mitigate problems from both effects in a classic global linear regression model [13]. GWR enables to estimate local parameter values for each region in the data set. The corresponding model is as follows [14]:

$$y_i = \beta_{i0} + \sum_{k=1}^{p-1} \beta_{ik}\, x_{ik} + \varepsilon_i \tag{1}$$

where index $i = 1, \ldots, n$, denotes the i-th region, y_i is the value of dependent variable at region i, x_{ik} denotes the values of the k-th independent variable at region i, β_{i0} is the intercept, β_{ik} is the regression parameter for the k-th independent variable, p is the number of regression terms, and ε_i denotes the error term at region i.

The local regression parameters are functions of region i and can be estimated by the weighted least squares for each region based on data from all the regions falling within a specified neighbourhood [8], [13], [14]:

$$\widehat{\boldsymbol{\beta}_i} = (\boldsymbol{X}^T \boldsymbol{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}_i \boldsymbol{y} \tag{2}$$

where $\boldsymbol{\beta_i}$ is the vector of p local regression parameters at region i, $\boldsymbol{y}$ is the $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ denotes the $n \times k$ matrix of independent variables (including a column of ones for the intercept), $\boldsymbol{W_i}$ is the $n \times n$ diagonal local weight matrix at region i whose off-diagonal elements are zero and diagonal elements are calculated based on a spatial kernel function giving higher weight to the nearby regions in comparison to regions farther away ([7], [14]). Matrix $\boldsymbol{W_i}$ is as follows:

$$\boldsymbol{W_i} = \left(\begin{bmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{bmatrix} \right)$$

where w_{ij} $(j = 1,2, \ldots, n)$ denote weights for each of the n observations for the regression point i.

The first step of the GWR approach is the choice of the kernel function and its bandwidth parameter. We can distinguish two types of kernel functions, fixed and adaptive. While the fixed kernel function is based on the same spatial range in each local kernel, the adaptive kernel function is characterized by the same number of observations in each local kernel. Fotheringham, Brunsdon and Charlton [7] and Wheeler and Páez [14] present various possibilities how to determine the optimal value of the bandwidth, e.g., minimising a cross validation score – *CV* or the corrected Akaike Information Criterion – *AICc*. Mapping of the estimated local parameters and/or of corresponding *t*-values allow us to uncover something which is hidden in the global regression analysis and thus to assess the stationarity or non-stationarity of analysed relationships [8]. In order to test whether the GWR local model (1) describes the data significantly better than a classic global linear regression model, the GWR ANOVA test can be used. Further testing procedures testing e.g. the spatial variation of the estimated local regression parameters can be found in Leung, Mei and Zhang [10].

3 Data and Empirical Results

The empirical part of the paper is based on the data of T5 average percentage Maths scores for the 79 Slovak districts retrieved from the web-site of NÚCEM[3] [15] for the school year 2018/2019. The shape file of the Slovak districts was downloaded from the web-site [16]. In order to analyse the impact of socioeconomic variables on the districts' school performance (measured by T5 average percentage Maths scores) the independent variables[4] – average nominal monthly wage, unemployment rate, and index of economic dependence of young people[5] in a district were downloaded from the DATAcube database of the Statistical Office of the Slovak Republic [17]. Analyses were performed in the free downloadable softwares GeoDa and GWR4.

Figure 1 illustrates the percentile maps for T5 average percentage Maths scores (denoted as *apm*), average nominal monthly wage (*w17*), unemployment rate (*un17*) and index of economic dependence of young people (*iy17*) across individual Slovak districts in order to visualise the unequally distribution of analysed variables over space. The best percentage Maths scores (73.8%) were reached by the pupil's in the district Bratislava I, high values between 66.76% and 72.785% were detected also in districts Bratislava II – Bratislava V, Košice I and Žilina. The least successful results were reached by the pupils in Gelnica district (38.0%), less successful (38.087% – 46.16%) were also pupils in districts of Kežmarok, Košice-okolie, Revúca, Rimavská Sobota, Rožňava, Trebišov and Vranov nad Topľou. Very large regional differences were confirmed also regarding the selected independent socioeconomic variables. The highest average nominal monthly wages were in Bratislava IV district followed by remaining districts of Bratislava, Košice II, Kysucké Nové Mesto and Trnava whereas the lowest values were identified for Banská Štiavnica district and seven other districts mainly in the eastern and south-eastern part of Slovakia. Enormous disparities are also observable for the unemployment rate. The difference between unemployment rate in the district of Trnava with the lowest rate (1.98%) and the district of Rimavská Sobota (18.48%) is about 16.5 percentage points. Mapping of the last independent variable, index of economic dependence of young people, indicates that the highest amount of children (0 – 14 years old) per 100 persons aged 15 to 64 was in Kežmarok district (35.72), high ratios of children to the population in productive age was detected also in five other districts in eastern part of Slovakia, in Námestovo district and district of Senec as well. On the other hand, the Myjava district reached considerably low values indicating only 18.19 children per 100 persons in productive age (i.e. aged 15 – 64).

The visual analysis is followed by OLS estimation of the global linear regression model assuming the consistent relationship between the average percentage Maths scores and the selected socioeconomic variables across the all Slovak districts – see Table 1 (column: Linear model OLS). All the estimated parameters were statistically significant at the 5 percent level of significance indicating the positive impact of average nominal monthly wages and negative impact of the remaining two independent variables, i.e. of unemployment rate and of economic dependence of young people on the analysed Maths scores. To investigate the spatial non-stationarity across analysed regions, i.e. spatially varied effects of selected socioeconomic variables on the average percentage Maths scores, the local GWR analysis was provided (the adaptive bi-square kernel based on 49 nearest neighbours was used). Selected GWR estimation results are presented in Table 1 (columns: GWR).

The estimated set of local parameters from the GWR fit together with the local coefficients of determination R_i^2 is mapped in Figure 2 which clearly documents the different model performance in individual regions. Although the average nominal monthly wage shows a globally significant positive effect on average percentage Maths scores, there are several districts in the eastern part of Slovakia, where the effect was negative. The impact of the second socioeconomic variable, the unemployment rate, was negative both in the global model fit as well as in the local GWR fit across all analysed regions. Since the global regression indicates the statistically significant negative impact of the last variable, index of economic dependence of young people, neither clear negative nor clear positive effect of these variable was proved for the GWR fit. Whereas the negative impact was proved for the regions in the eastern and middle part of Slovakia, in the western and southern regions it was positive. The signs of the local parameters were in line with those of global model for 44 districts (in the western and middle part of Slovakia) out of the 79 Slovak districts indicating the positive impact of the average nominal monthly wage and negative impact of both the unemployment rate and index of economic dependence of young people. In the remaining districts were the local relationships different.

[3] Testing 5, i.e. external testing of pupils of the 5th year of primary schools, has been performed by the National Institute for Certified Educational Measurements (Národný ústav certifikovaných meraní vzdelávania „NÚCEM") since the school year 2015/2016 in order to monitor pupils' level of knowledge and skills and to obtain relevant information about their performance at the entrance to the 2nd stage of primary school.

[4] Regarding the data availability, we used the values related to the year 2017.

[5] Index of economic dependence of young people = the number of persons 0 – 14 years old per 100 persons aged 15 to 64.

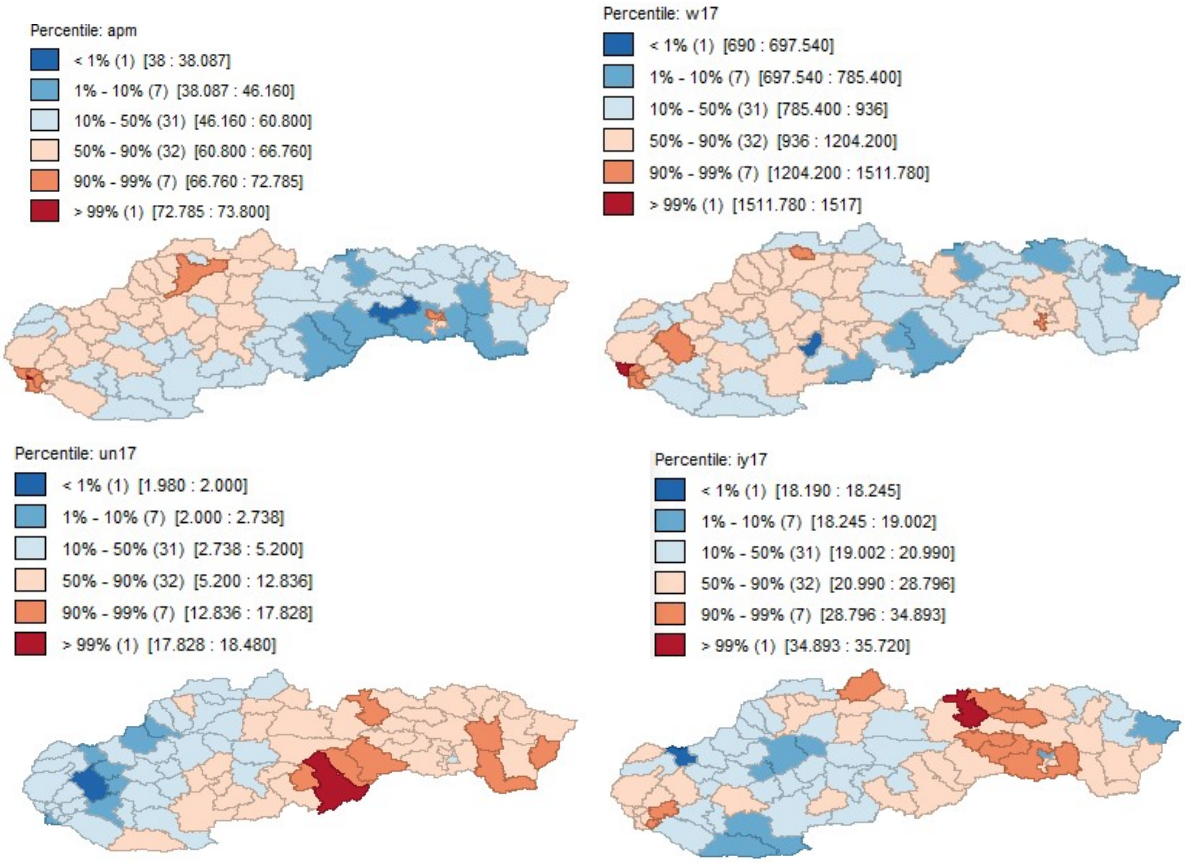


Figure 1 Percentile maps for T5 average percentage Maths scores (*apm*), average nominal monthly wage (*w17*), unemployment rate (*un17*) and index of economic dependence of young people (*iy17*)

Model	**Linear model**	**GWR**				
	OLS	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
β_0	62.327	41.242	43.143	59.646	81.506	91.098
β_1 (*w17*)	0.011	-0.004	0.002	0.007	0.013	0.014
β_2 (*un17*)	-1.272	-1.773	-1.650	-1.249	-0.489	-0.328
β_3 (*iy17*)	-0.255	-0.682	-0.545	0.114	0.334	0.414
AICc	438.832			409.325		
Adjusted R^2	0.767			0.853		

Table 1 Estimation results of OLS regression and of GWR

The goodness-of-fit based on values of the adjusted R^2 and corrected Akaike Information Criterion (*AICc*) for the global linear regression model and local GWR model, respectively, indicate the clear evidence of improvement in the model performance. The presented results showed that there is a strong relationship between the regional school performance and the socioeconomic variables of the region. Since the residuals from the global linear regression OLS model fit showed statistically significant positive spatial autocorrelation with Moran's *I* of 0.285, in the residuals from the GWR fit there was no evidence of spatial autocorrelation (Moran's *I* of 0.062)[6]. The statistically significant improvement in the GWR model performance over the global linear regression model was confirmed by the GWR ANOVA test with the test statistic $F = 5.683$. Using the geographical variability tests of

[6] The Moran's *I* values were calculated based on the queen case definition of spatial weights. For more information about testing for spatial autocorrelation see e.g., [2].

local parameters it was confirmed that all parameters exhibited statistically significant spatial variations in their values. Our results are in line with those of [4], [8], [11] confirming that the GWR analysis enables to uncover deeper patterns hidden in the global analysis.

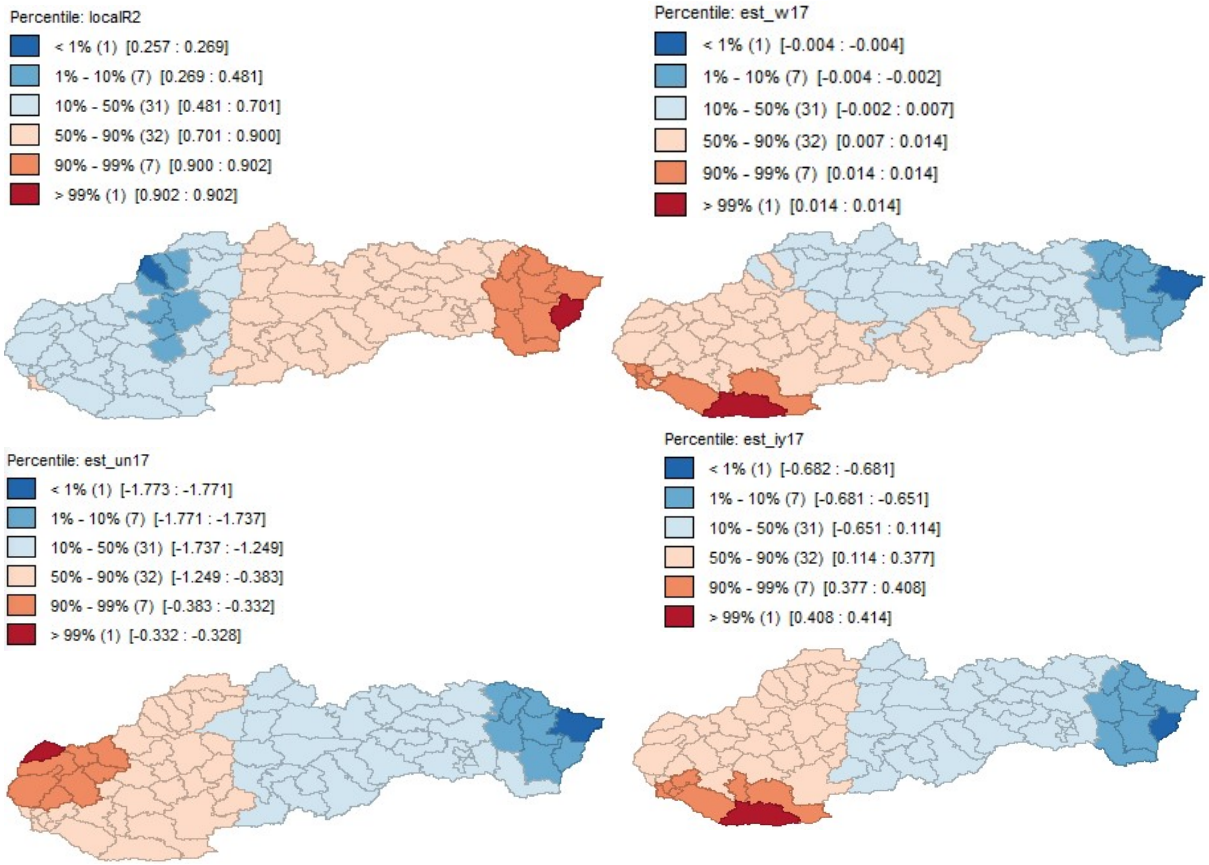


Figure 2 Local coefficients of determination R_i^2 and local parameters of average nominal monthly wage (*w17*), unemployment rate (*un17*) and index of economic dependence of young people *(iy17)* from the GWR fit

4 Conclusion

The spatial analysis of school performance as well as studying of various socioeconomic factors affecting it is a challenging social and political issue. Based on the T5 average percentage Maths scores huge differences were recorded across individual Slovak districts. The paper analyses the impact of selected socioeconomic indicators (average nominal monthly wage, unemployment rate and index of economic dependence of young people) characterizing the environment of individual districts on the test scores achievement. The global regression results for Slovakia as a whole confirmed (as expected) the positive impact of wage on the test results and negative impact of unemployment rate and of index of economic dependence of young people. In order to capture the spatial variation in the modelled relationship, the local GWR approach enabled to reveal quite a high amount of districts, located especially in the middle and eastern part of Slovakia, with locally different impacts of analysed socioeconomic indicators.

To analyse the test scores of Testing 9 as well as to consider the impact of further socioeconomic variables and mapping the results of the local GWR approach could be the areas of the future research.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA 1/0248/17 "Analysis of Regional Disparities in the EU based on Spatial Econometric Approaches".

References

[1] Anselin, L. (2001). *Spatial Econometrics. – Companion to Econometrics*. In: Baltagi, B. (Ed.), Oxford: Basil Blackwell, 2001.

[2] Arbia, G. (2006). *Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence*. Berlin Heidelberg: Springer-Verlag.

[3] Chocholatá, M. (2018). Spatial Analysis of Convergence in the Selected Regions of Central Europe. In *Proceedings of the 18th International Scientific Conference „AIESA – BUILDING OF SOCIETY BASED ON KNOWLEDGE"(pp. 19-30)*. Bratislava: Letra Edu.

[4] Christie, P. (2013). Space, Place, and Social Justice: Developing a Rhythmanalysis of Education in South Africa. *Qualitative Inquiry*, ***19***(10), 775-785.

[5] Conduit, E., Brooks, R., Bramley, G. & Fletcher, C.L. (1996). The value of school locations. *British Educational Research Journal*, ***22***, 199-206.

[6] Coombes, M. & Raybould, S. (1997). Modelling the influence of individual and spatial variations in the levels of secondary school examination results. *Environment and Planning A*, ***29***, 641-658.

[7] Fotheringham, A. S., Brunsdon, C. & Charlton, M. E. (2002). *Geographically Weighted Regression. The Analysis of Spatial Varying Relationships*. Chichester: Wiley.

[8] Fotheringham, A.S., Charlton, M.E. & Brunsdon, C. (2001). Spatial Variations in School Performance: a Local Analysis Using Geographically Weighted Regression. *Geographical & Environmental Modelling*, ***5***(1), 43-66.

[9] Furková, A. (2018). Spatial Heterogeneity of Regional Innovation Processes: Geographically Weighted Regression Approach. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics: Multiple Criteria Decision Making XIX (pp. 127-134).* Bratislava: Letra Edu.

[10] Leung, Y., Mei, C., & Zhang, W. (2000). Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environment and Planning A*, ***32***(1), 9-32.

[11] Naidoo, A.G.V., van Eeden, A. & Munch, Z. (2014). Spatial Variation in School Performance, a Local Analysis of Socio-economic Factors in Cape Town. *South African Journal of Geomatics, **3***(1)*,* 78-94.

[12] Národný ústav certifikovaných meraní vzdelávania. (2019). Testovanie 5 – 2018 – Výsledky celoslovenského testovania žiakov 5. ročníka ZŠ (Tlačová správa). Available online at: https://www.nucem.sk/dl/4086/Tla%C4%8Dov%C3%A1%20spr%C3%A1va_T5_2018.pdf [cited 2019-02-07].

[13] Qiu, X. & Wu, S.-S. (2011). Global and Local Regression Analysis of Factors of American College Test (ACT) Score for Public High Schools in the State of Missouri. *Annals of the Association of American Geographers*, ***101***(1), 63-83.

[14] Wheeler, D.C., & Páez, A. (2010). Geographically Weighted Regression. In Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications (pp. 461-486). Berlin Heidelberg: Springer-Verlag.

[15] www.nucem.sk [cited 2019-02-07].

[16] http://wiki.freemap.sk/HraniceAdministrativnychUzemi [cited 2015-02-11].

[17] http://datacube.statistics.sk/ [cited 2019-02-07].

Dynamic efficiency analysis of German NUTS 2 regions

Petra Zýková[1]

Abstract. The paper deals with efficiency analysis of 38 German NUTS 2 (Nomenclature of Territorial Units for Statistics) regions. The aim is to find the most efficient regions and their ranking between the years 2008 – 2016. The efficiency analysis is based on an application of data envelopment analysis (DEA) models. This paper uses dynamic DEA models window analysis that is one of the tools for dealing with time factor in DEA. The data set contains information about the NUTS 2 regions for nine following years starting by 2008. There are used two inputs – employment (thousand hours worked) and gross fixed capital formation (million Euro) and one output – gross domestic product (million Euro).

Keywords: data envelopment analysis, time series, window analysis, efficiency, Germany
JEL Classification: C44

AMS Classification: 90C05, 90C90

1 Introduction

Often, we need to compare single districts of countries. Many ways may solve this task. One of the possibilities is to apply data envelopment analysis (DEA) models. DEA models have been first developed by Charnes et al. [2] based on the concept introduced by Farrell [4]. Using DEA models, the efficiency scores of decision-making units are computed. Basic DEA models do not deal with time series, due to dynamic DEA models are proposed. Development of efficiency score in time is a significant factor. This paper uses window analysis based on moving averages for dealing with time shown in [3]. After that, the comparison between dynamic and static analysis is set. Two static analyses are used: analysis for particular years and overall analysis. This paper compares concretely 38 NUTS 2 regions in Germany. Efficiency analysis of NUTS 2 regions have been investigated by [5] by Malmquist index and in [1].

The paper is organized as follows. The next section presents the definition of all models used in the study, i.e. DEA models and dynamic DEA models that takes into account time series of inputs and/or outputs. Section 3 contains a dynamic and static analysis of German NUTS 2 regions and comparison between used models. The last section of the paper concludes the results and discusses future research.

2 DEA models and dynamic analysis

DEA models are a general tool for efficiency and performance evaluation of the set of homogenous DMUs that spend multiple w inputs and transform them into multiple t outputs. The measure of efficiency (efficiency score) of this transformation is one of the main results of the application of DEA models. Let us denote $\mathbf{Y} = \left(y_{rj},\ r = 1,\dots,t,\ j = 1,\dots n\right)$ a non-negative matrix of outputs and $\mathbf{X} = \left(x_{kj},\ k = 1,\dots,w, j = 1,\dots n\right)$ a non-negative matrix of inputs. The efficiency score of the unit under evaluation DMU_{j_0} is derived as follows:

Maximize

$$U_{j_0} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{k=1}^{w} v_k x_{kj_0}}$$

subject to

$$\begin{aligned} &\frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{k=1}^{w} v_k x_{kj}} \le 1, \quad j = 1,\dots,n, \\ &u_r \ge \varepsilon, \quad r = 1,\dots,t, \\ &v_k \ge \varepsilon, \quad k = 1,\dots,w, \end{aligned} \tag{1}$$

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

where u_r is a positive weight of the r^{th} output, v_k is a positive weight of the k^{th} input, and ε is an infinitesimal constant. Model (1) is not linear in its objective function but may easily be transformed into a linear program. The linearized version of the input-oriented model (often called the CCR model) is as follows:
Maximize

$$U_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

$$\sum_{k=1}^{w} v_k x_{kj_0} = 1,$$

subject to

$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{k=1}^{w} v_k x_{kj} \le 0, \quad j = 1,\dots,n, \tag{2}$$

$$u_r \ge \varepsilon, \quad r = 1,\dots,t,$$

$$v_k \ge \varepsilon, \quad k = 1,\dots,w.$$

Above mentioned DEA models analyse efficiency score of DMU only in one time period. The simplest way how to cover the development of efficiency score in time is window analysis. This dynamic DEA method of efficiency analysis is based on moving averages. In this method, it is approached to a certain DMU in two different time periods as two different independent units in one time. We have a set of n DMUs, which are defined by w inputs and t outputs in τ, $\tau = 1,\dots,T$ time periods. The windows are overlapping. Depending on the size of the window p is the total number of windows $z = T - p + 1$. That is why, for every DMU under evaluation we get $p(T - p + 1)$ efficiency scores. From those efficiency scores, we count the average

$$E_{j_0} = \frac{\sum_{b=1}^{z}\sum_{c=1}^{p} U_{bj_0}^{c}}{z \cdot p}, \quad j_0 = 1,\dots,n, \tag{3}$$

where $U_{bj_0}^{c}$ is the efficiency score of j_0 th DMU in the b th window and in this window in the time c.

3 The case of Germany NUTS 2 regions

Germany (Bundesrepublik Deutschland) is consisting of 16 federal countries (Bundesland). Federal countries are NUTS 1. Further, Germany is divided into 38 regions (Bezirk) level NUTS 2. In this analysis, we are investigating efficiency scores of these NUTS 2 regions for nine following years from the year 2008 to 2016. We have used these inputs: labour - employment in thousands of hours and capital – gross fixed capital in millions of euros. And one output: GDP in millions of euros. The data was obtained from [6].

The number of DMUs is $n = 38$, the number of inputs $w = 2$, the number of outputs $t = 1$, and the number of time periods $T = 9$.

3.1 Dynamic analysis – window analysis

We have set the size of window $p = 3$ for this analysis, and so, the total number of windows is $z = 7$. There are $n \cdot p = 38 \cdot 3 = 114$ DMUs in every window. Thus, Model (2) was applied for every window seven times, always for the different dataset. The efficient units (region/year) in particular windows are presented in Table 1.

window	NUTS 2 region	year
1	Hamburg	2008
	Bremen	2009
	Hamburg	2009
	Darmstadt	2010
	Düsseldorf	2010
2	Bremen	2009
	Hamburg	2009
	Darmstadt	2010
	Düsseldorf	2010
	Hamburg	2011
3	Düsseldorf	2010
	Oberbayern	2011
	Hamburg	2012
	Darmstadt	2012
	Düsseldorf	2012
4	Hamburg	2013
	Darmstadt	2013
	Düsseldorf	2013
5	Bremen	2014
	Hamburg	2014
	Darmstadt	2014
6	Bremen	2014
	Hamburg	2015
	Darmstadt	2015
7	Bremen	2014
	Darmstadt	2015
	Hamburg	2016
	Düsseldorf	2016

Table 1 Efficient NUTS 2 regions in particular windows.

There have been found 28 efficient NUTS 2 regions in Table 1. Hamburg was determined most often as the efficient region, nine times and at least once in every particular window – see Table 2.

NUTS 2 region	efficient
Hamburg	9
Darmstadt	7
Düsseldorf	6
Bremen	5
Oberbayern	1

Table 2 How many times was efficient NUTS 2 region in single windows.

For every region $p(T-p+1)=3(9-3+1)=21$ efficiency scores were computed. According to (3) was computed the average for every particular NUTS 2 region. There are averages in Table 3 for the first ten most efficient NUTS 2 regions.

NUTS 2 region	average	ranking
Düsseldorf	0.988693	1
Darmstadt	0.988346	2
Hamburg	0.987742	3
Bremen	0.964551	4
Köln	0.954420	5
Oberbayern	0.954144	6
Arnsberg	0.929282	7
Stuttgart	0.918269	8
Hannover	0.876437	9
Karlsruhe	0.875369	10

Table 3 Average efficiency scores and ranking of ten most efficient NUTS 2 regions.

There are shown average efficiency scores for eight the less efficient NUTS 2 regions in Table 4.

NUTS 2 region	average	ranking
Sachsen-Anhalt	0.743807	31
Leipzig	0.721729	32
Chemnitz	0.714484	33
Thüringen	0.702490	34
Brandenburg	0.700364	35
Trier	0.692924	36
Mecklenburg-Vorpommern	0.658876	37
Dresden	0.654308	38

Table 4 Average efficiency scores and ranking of eight less efficient NUTS 2 regions.

3.2 Static analysis

The impact of time is not covered in the static analysis. We could approach dataset in two ways. Firstly, we compute model (2) for every year separately and then compute the average of obtained efficiency scores and the ranking of NUTS 2 regions. Or secondly, we can consider NUTS 2 regions as independent of time and apply model (2) for $n \cdot T = 38 \cdot 9 = 342$ independent units.

year	2008	2009	2010	2011	2012	2013	2014	2015	2016	average	ranking
Hamburg	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1 - 2
Darmstadt	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1 - 2
Düsseldorf	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9837	1.0000	1.0000	0.9982	3
Bremen	0.9686	1.0000	0.9768	1.0000	0.9148	0.9542	1.0000	0.9472	0.9810	0.9714	4
Oberbayern	0.9324	0.9206	0.9487	0.9870	1.0000	0.9831	0.9872	0.9778	0.9915	0.9698	5
Köln	0.9431	0.9655	0.9744	0.9553	0.9793	0.9712	0.9544	0.9739	0.9804	0.9664	6
Arnsberg	0.9133	0.9027	0.9119	0.9208	1.0000	0.9296	0.9536	0.9652	0.9724	0.9411	7
Stuttgart	0.8975	0.8730	0.9417	0.9523	0.9425	0.9248	0.9403	0.9587	0.9473	0.9309	8
Hannover	0.9424	0.8727	0.8849	0.9000	0.8831	0.8780	0.8784	0.8840	0.8976	0.8912	9
Karlsruhe	0.8792	0.8652	0.9014	0.9127	0.8752	0.8951	0.8757	0.8796	0.9001	0.8871	10

Table 5 Efficiency scores for particular years, average and ranking of the ten most efficient NUTS 2 regions.

Analysis for single years

Model (2) is solved nine times (for each year) in this case. There are efficiency scores for particular years and average efficiency score for all nine years for ten most efficient NUTS 2 regions presented in Table 5. In this table we can see that regions Hamburg and Darmstadt are efficient in every single investigated year. Düsseldorf took third place in this analysis. Düusefdorf was efficient in seven from nine investigated years. The fourth most efficient region is region Bremen, which was efficient third times. Regions Oberbayern and Arsberg are worth mentioning, there were efficient in one year only.

Overall analysis

The impact of time is not covered in the analysis overall years. We approach all units independently. We have analysed 342 independent units. This solution gave us five efficient units in Table 6.

NUTS 2 region	year
Bremen	2009
Bremen	2014
Darmstadt	2015
Hamburg	2016
Düsseldorf	2016

Table 6 Efficient NUTS 2 regions from all regions in all years.

It is shown in Table 6 that region Bremen is efficient twice, in the years 2009 and 2014. Hamburg, Darmstadt, Düsseldorf were efficient once. These regions were determined as efficient in comparison with other regions.

NUTS 2 region	average	ranking
Düsseldorf	0.976885	1
Bremen	0.954339	2
Darmstadt	0.953818	3
Hamburg	0.938090	4
Arnsberg	0.924631	5
Köln	0.923260	6
Oberbayern	0.898376	7
Stuttgart	0.875761	8
Hannover	0.862981	9
Saarland	0.855921	10

Table 7 Average efficiency scores of NUTS 2 regions and ranking for the ten most efficient regions in all years.

There are shown average efficiency scores for ten average most efficient NUTS 2 regions computed by overall analysis in Table 7. The highest average has NUTS 2 region Düsseldorf, then Bremen and Darmstadt. Hamburg has the fourth highest average efficiency score.

3.3 Comparison between models

The average efficiency scores and ranking for ten most efficient NUTS 2 regions are shown in Table 8. In an application of dynamic windows analysis region Düsseldorf, which was six times efficient in the window analysis, was as the most efficient region determined. Düsseldorf took third place in the analysis of single years and in the overall analysis was again identified as the most efficient region.

In an application of dynamic windows analysis and then from those computed average efficiency score, was as the second most efficient region determined region Darmstadt, which was seven times efficient in the window analysis. Darmstadt took the first place together with Hamburg in the analysis of single years and in the overall analysis this region took the third place.

In an application of dynamic windows analysis and then from those computed average efficiency score, as the third most efficient region was determined Hamburg, which was nine times efficient in the window analysis.

Hamburg took the first place together with Darmstadt in the analysis of single years and took the fourth place in the overall analysis.

NUTS 2 region Bremen took the fourth place according to average efficiency scores obtained by dynamic window analysis and also according to the analysis of single years. Bremen took second place in the overall analysis. Bremen was efficient five times in the window analysis.

From the comparison between models follows that the fourth most efficient regions come out as the most efficient by all models mentioned above. If we want to only divide regions into "better" and "worse" regions, it does not matter which method we use. Nevertheless, for detailed, analysis of the most efficient regions is proper to use dynamic analysis. Personally, we determine the most efficient region according to Table 2, i.e. how many times was the region efficient in the particular windows. Therefore is the most efficient NUTS 2 region Hamburg, which was efficient nine times. Then on the second place Darmstadt, which was efficient seven times and then on the third pace Düsseldorf which was efficient six times.

	window analysis		**particular years**		**overall analysis**	
NUTS 2 region	**average**	**ranking**	**average**	**ranking**	**average**	**ranking**
Düsseldorf	0.9887	1	0.9982	3	0.9769	1
Darmstadt	0.9883	2	1.0000	1 - 2	0.9538	3
Hamburg	0.9877	3	1.0000	1 - 2	0.9381	4
Bremen	0.9646	4	0.9714	4	0.9543	2
Köln	0.9544	5	0.9664	6	0.9233	6
Oberbayern	0.9541	6	0.9698	5	0.8984	7
Arnsberg	0.9293	7	0.9411	7	0.9246	5
Stuttgart	0.9183	8	0.9309	8	0.8758	8
Hannover	0.8764	9	0.8912	9	0.8630	9
Karlsruhe	0.8754	10	0.8871	10	0.8456	12

Table 8 The most efficient NUTS 2 regions.

4 Conclusions

This paper dealt with an application of data envelopment analysis models for efficiency evaluation of NUTS 2 regions in Germany. There were used two inputs (labour and capital), and one output (GDP) in the CCR input oriented models. There were used dynamic window analysis and static analysis: analysis for particular years and overall analysis for all 38 NUTS 2 regions. Four most efficient regions have been identified. There are Bremen, Darmstadt, Düsseldorf and Hamburg. The results of all applied models are described in detail in the section above. The most efficient NUTS 2 region is Hamburg according to those analyses.

The numerical experiments were realized using original procedures written in the LINGO modelling language. Dynamic DEA models are an interesting subject for further research. In our future research, we plan to extend the influence of time in DEA models to use something more scientific than moving averages.

Acknowledgements

This work was supported by the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague, project F4/20/2018 (*Data envelopment analysis models in economic decision making*).

References

[1] Aristovnik, A. (2014). Efficiency of the R&D Sector in the EU-27 at the Regional Level: An Application of DEA. *Lex Localis – Journal of Local Self-Government*. 12(3), 519-531.
[2] Charnes, A., Cooper, W. and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp. 429-444.
[3] Dlouhý, M., Jablonský, J. and Zýková P. (2018) *Analýza obalu dat*. Praha: Professional Publishing.
[4] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3): 253-290.
[5] Staníčková, M., Melecký, L. (2016). Malmquist index approach to efficiency analysis in selected old and new EU Member States. *Ekonomická revue* 19(3), 87-104
[6] https://ec.europa.eu/eurostat/data/database

A Note on Method of Moments Estimation of Ornstein–Uhlenbeck Process Using Ultra-High-Frequency Data

Vladimír Holý[1]

Abstract. Stock prices, foreign exchange rates and commodity prices are recorded with each transaction or bid/ask offer resulting in intraday ultra-high-frequency data. Such time series have a very fine time scale and several distinctive characteristics including irregularly spaced observations and the presence of the market microstructure noise. When time series also exhibit mean-reverting behavior, the Ornstein–Uhlenbeck process with continuous values and continuous time can be used to model them. We propose an estimator of the Ornstein–Uhlenbeck process for irregularly spaced observations contaminated by the independent white noise. The estimator is based on the method of moments and utilizes the sample mean, the sample variance and the autocovariance function approximated by the least squares method. The advantage of the method of moments is that it does not require the underlying distribution to be specified. In a simulation study, we compare the proposed method of moments estimator with the maximum likelihood estimator.

Keywords: Ornstein–Uhlenbeck Process, Ultra-High-Frequency Data, Market Microstructure Noise, Method of Moments.

JEL Classification: C22, C58
AMS Classification: 60J60, 62F10

1 Introduction

In finance, *ultra-high-frequency data* refer to irregularly spaced time series recorded at the highest possible frequency corresponding to each transaction or change in bid/ask offer [5]. Financial high-frequency time series include stock prices, foreign exchange rates and commodity prices. The availability of these high-frequency data allows econometricians to construct more precise models while facing new challenges. One of the distinctive features of high-frequency data is the presence of the *market microstructure noise*. It is caused by the bid-ask bounce, discretness of price values and various informational effects and conceals the true volatility of the price process [7]. For an extensive overview of financial high-frequency data analysis, see [1], [8] and [9].

Many financial time series tend to move to their mean values over time. This behavior is known as the *mean reversion* and is often captured by the *Ornstein–Uhlenbeck process* [14]. In finance, the Ornstein–Uhlenbeck process can be used to model currency exchange rates [2], commodity prices [12], interest rates [15], spread between correlated assets [4] and stochastic volatility [3].

The paper [10] deals with estimation of the Ornstein–Uhlenbeck process in the high-frequency setting. Specifically, the presence of the market microstructure noise and irregularly spaced observations are considered. Three estimators robust to the noise are proposed in the paper – the method of moments for regularly spaced observations, the approach based on ARMA(1,1) reparametrization for regularly spaced observations and the maximum likelihood estimator for irregularly spaced observations. The empirical study of 7 Big Oil companies shows that the noise-robust approach leads to a significant increase in profitability of the pairs trading strategy.

We follow [10] and propose a novel noise-robust estimator based on the method of moments for irregularly spaced observations. Generally, the method of moments does not provide sufficient statistic nor asymptotically efficient estimator and is superseded by the maximum likelihood estimator [6]. However, the method of moments does not require the underlying distribution to be specified. In contrast, the maximum likelihood estimator of [10] assumes Gaussianity for both the distribution driving the Ornstein–Uhlenbeck process and the distribution of the market microstructure noise. We compare both methods in a simulation study under various noise scenarios.

The rest of this paper is structured as follows. In Section 2, we introduce the Ornstein–Uhlenbeck process in the ultra-high-frequency setting. In Section 2, we propose the method of moments estimator. In Section 4, we conduct the simulation study. We conclude the paper in Section 5.

[1] University of Economics, Prague, vladimir.holy@vse.cz

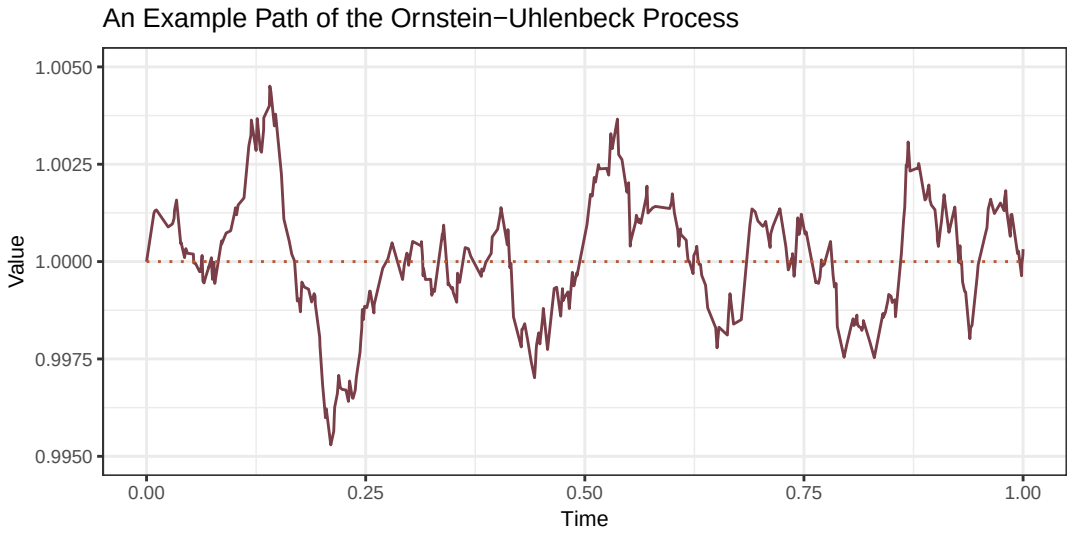


Figure 1 A simulated path of the Ornstein–Uhlenbeck process with parameters $\mu = 1$, $\tau = 10$ and $\sigma^2 = 10^{-4}$.

2 Theoretical Framework

We consider the same high-frequency framework for the Ornstein–Uhlenbeck process estimation as in [10]. The *Ornstein–Uhlenbeck process* $P_t, t \geq 0$ is a process satisfying stochastic differential equation

$$\mathrm{d}P_t = \tau(\mu - P_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{1}$$

where W_t is the Wiener process, μ is a parameter representing *long-term mean*, $\tau > 0$ is a parameter representing *speed of reversion* and $\sigma > 0$ is a parameter representing *instantaneous volatility*. This stochastic differential equation has solution

$$P_t = P_0 e^{-\tau t} + \mu(1 - e^{-\tau t}) + \sigma \int_0^t e^{-\tau(t-s)} \mathrm{d}W_s. \tag{2}$$

An example path of the process is illustrated in Figure 1.

The Ornstein–Uhlenbeck process is defined in continuous time. However, we do not observe continuous paths of the process in practice. Instead, we only observe the process P_{T_i} at a finite number of times $T_0 < T_1 < \ldots < T_n$. We consider times of observations T_i to be deterministic. Without loss of generality, we restrict ourselves to time interval $[0, 1]$, i.e. $T_0 = 0$ and $T_n = 1$. We further assume that the observed process is contaminated by the independent white noise with zero mean and variance ω^2. For the observed discrete process X_i, we utilize the *additive noise model*

$$X_i = P_{T_i} + E_i, \qquad i = 0, \ldots, n, \tag{3}$$

where $E_i \overset{i.i.d.}{\sim} (0, \omega^2)$ is the so-called *market microstructure noise*. In contrast with paper [10] which considers only Gaussian noise, we allow for any distribution with zero mean and finite variance.

3 Method of Moments

The *method of moments* is based on relating theoretical values of moments to their finite-sample estimates. We utilize the expected value, the variance and the autocovariation function. The moments of the observed process X_i are given by

$$\begin{aligned} \mathrm{E}[X_i] &= \mu, \\ \mathrm{var}[X_i] &= \frac{\sigma^2}{2\tau} + \omega^2, \\ \mathrm{cov}[X_i, X_j] &= \frac{\sigma^2}{2\tau} e^{-\tau|T_i - T_j|}, \qquad i \neq j. \end{aligned} \tag{4}$$

The autocovariance function is illustrated in Figure 2. Note that the variance of the observed process (the case $|T_i - T_j| = 0$) is increased by the noise variance ω^2.

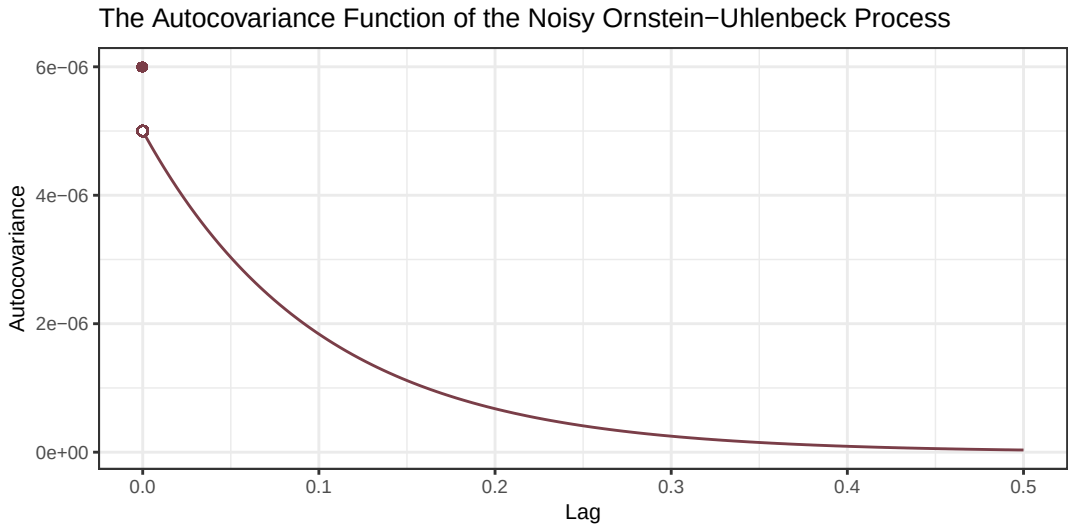


Figure 2 The autocovariance function of the Ornstein–Uhlenbeck process contaminated by the white noise with parameters $\mu = 1$, $\tau = 10$, $\sigma^2 = 10^{-4}$ and $\omega^2 = 10^{-6}$.

We estimate the first moment using observed values $x_0, x_1, \ldots, x_n$ as

$$\hat{m} = \frac{1}{n+1} \sum_{i=0}^{n} x_i. \tag{5}$$

The long-term mean parameter μ is then simply estimated as

$$\hat{\mu} = \hat{m}. \tag{6}$$

We approximate the autocovariance function by the *least squares* method. The summands in sample covariance $(x_i - \hat{m})(x_j - \hat{m})$ act as the dependent variable while the time lags $|T_i - T_j|$ are the independent variable. The regression model is nonlinear and has the form

$$(x_i - \hat{m})(x_j - \hat{m}) = \frac{\sigma^2}{2\tau} e^{-\tau |T_i - T_j|} + \varepsilon_{i,j}, \qquad 0 \le i < j \le \min\{i + k, n\}, \tag{7}$$

where $\varepsilon_{i,j}$ are the error terms and $k \in \mathbb{N}^+$ denotes the maximal considered lag. Theoretically, the maximal lag k can be equal to n and therefore not restrictive. In practice, however, large values of k lead to a huge number of observations as there are many combinations of i and j. We discuss computational issues regarding this parameter in Section 4. Parameters τ and σ^2 are then estimated as

$$\left(\hat{\tau}, \hat{\sigma}^2\right) = \underset{\tau, \sigma^2}{\arg\min} \sum_{\substack{0 \le i < j \le n, \\ j \le i + k}} \left((x_i - \hat{m})(x_j - \hat{m}) - \frac{\sigma^2}{2\tau} e^{-\tau |T_i - T_j|} \right)^2. \tag{8}$$

Finally, we estimate the variance of the process as

$$\hat{v} = \frac{1}{n} \sum_{i=0}^{n} (x_i - \hat{m})^2. \tag{9}$$

The variance of the noise ω^2 is then estimated as

$$\hat{\omega}^2 = \hat{v} - \frac{\hat{\sigma}^2}{2\hat{\tau}}. \tag{10}$$

A disadvantage of the method of moments is that it does not enforce parameter restrictions such as positivity of τ, σ^2 and ω^2. When the estimate is not positive, we simply set it to a very small positive threshold.

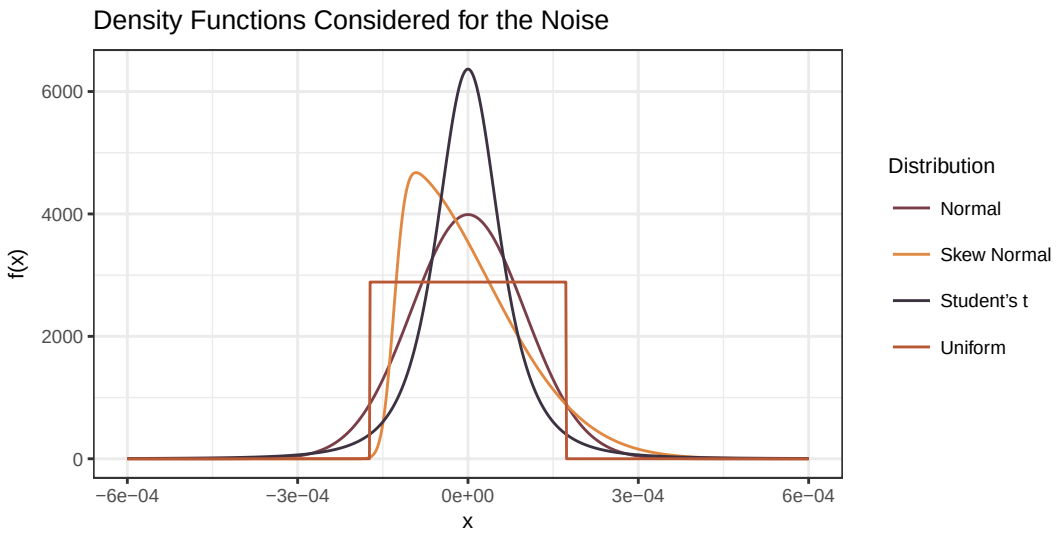


Figure 3 Density functions of the considered continuous distributions with zero mean and variance $\omega^2 = 10^{-8}$.

4 Simulation Study

We compare the proposed method of moments estimator with the maximum likelihood estimator of [10] (labeled as MLE). For the method of moments, we consider three values of the maximal lag – $k = 10$ (labeled as MOM-10), $k = 100$ (labeled as MOM-100) and $k = 1\,000$ (labeled as MOM-1000). The number of observations in each simulation is set to $n = 23\,400$ corresponding to durations between price changes to be one second on average during 6.5 hours long trading day. The number of combinations of i and j in (7) is then $\sum_{l=n-k+1}^{n} l$. Specifically, we have 233 955 combinations for $k = 10$, 2 335 050 combinations for $k = 100$ and 22 900 500 combinations for $k = 1\,000$. With a significantly larger maximal lag k, we would face memory issues during the computation.

We simulate the Ornstein–Uhlenbeck process with parameters $\mu = 10^{-1}$, $\tau = 10$ and $\sigma^2 = 10^{-4}$. We further contaminate the Ornstein–Uhlenbeck process by the independent white noise with zero mean and variance $\omega^2 = 10^{-8}$ using the additive model (3). The values of the parameteres are taken from the simulation study of [10] and resemble empirical values of stock prices of the 7 Big Oil companies. For the noise, we consider four continuous distributions – the *normal distribution*, the *skew normal distribution* [11] with the shape parameter equal to 10, the *Student's t-distribution* with 3 degrees of freedom and the *uniform distribution*. All continuous distributions are standardized to zero mean and variance ω^2. Density functions are illustrated in Figure 3. We also consider two discrete distributions – the *Rademacher distribution* (values 1 and -1, each with probability 0.5) and the *Skellam distribution* [13]. Both discrete distributions are standardized to zero mean and variance ω^2 with the support set to multiples of ω. Probability mass functions are illustrated in Figure 4. The simulated observations are irregularly spaced and the times of observations are generated by the *Poisson point process*. We perform the simulation 10 000 times for each of the 6 noise scenarios.

Results of the simulation study are reported in Table 1. For the comparison of parameter estimates, we utilize the *mean absolute error*. In each of the 6 scenarios, the method of moments has lower error for the long-term mean parameter μ. For the remaining three parameters τ, σ^2 and ω^2, the maximum likelihood method performs better.

The error of the method of moments estimates of τ and σ^2 decreases with higher maximal lag k. Therefore, we recommend to use as high k as computational limits allow. Interestingly, this does not hold for the noise variance ω^2. For all three values of k, the method of moments has very high error suggesting (10) is not very suitable for estimation of ω^2. This is caused by inaccurate estimates of the unconditional variance of the process. The sample variance often underestimates the theoretical variance and the resulting estimate of ω^2 is negative. Overall, the method of moments gives decent estimates of the Ornstein-Uhlenbeck process parameters μ, τ and σ^2 when enough lags between observations are utilized, but gives very poor estimates of the noise variance parameter ω^2.

The maximum likelihood estimator proved to be quite robust to misspecification of the noise distribution. It assumes normal distribution, but gives similarly precise estimates when the actual distribution is the skew normal, uniform, Rademacher or Skellam. Only the Student's t-distribution distinctly increases the error of parameter estimates suggesting distributions with heavy tails are more challenging when assuming Gaussianity.

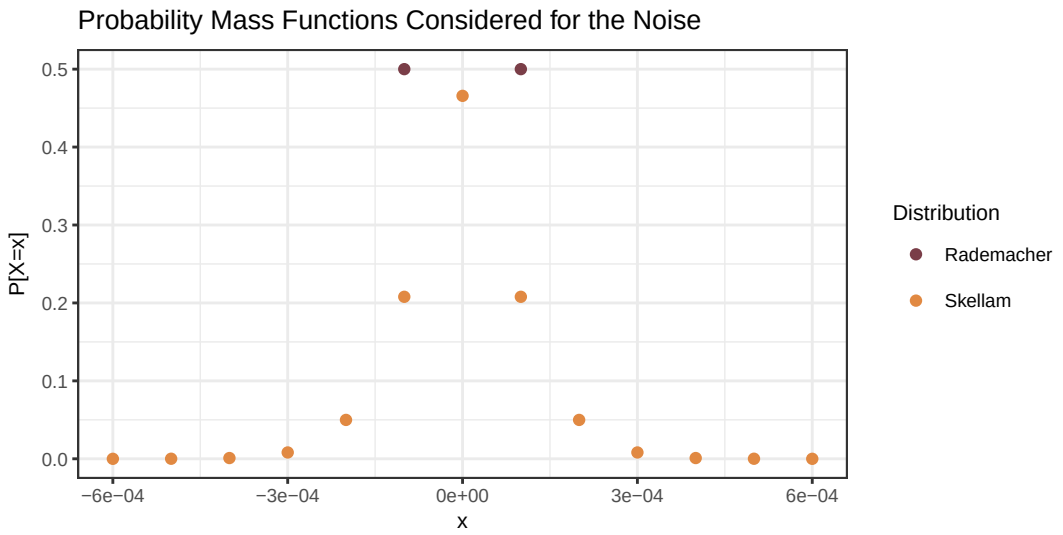


Figure 4 Probability mass functions of the considered discrete distributions with zero mean and variance $\omega^2 = 10^{-8}$.

Noise Distribution	Method	$\hat{\mu}$	$\hat{\tau}$	$\hat{\sigma}$	$\hat{\omega}$
Normal	MOM-10	$7.6051 \cdot 10^{-4}$	$0.3809 \cdot 10^{2}$	$1.1141 \cdot 10^{-2}$	$0.6861 \cdot 10^{-4}$
Normal	MOM-100	$7.6051 \cdot 10^{-4}$	$0.0991 \cdot 10^{2}$	$0.3718 \cdot 10^{-2}$	$1.2292 \cdot 10^{-4}$
Normal	MOM-1000	$7.6051 \cdot 10^{-4}$	$0.0748 \cdot 10^{2}$	$0.1418 \cdot 10^{-2}$	$1.3068 \cdot 10^{-4}$
Normal	MLE	$7.6219 \cdot 10^{-4}$	$0.0544 \cdot 10^{2}$	$0.0261 \cdot 10^{-2}$	$0.0066 \cdot 10^{-4}$
Skew Normal	MOM-10	$7.6214 \cdot 10^{-4}$	$0.3662 \cdot 10^{2}$	$1.0968 \cdot 10^{-2}$	$0.6770 \cdot 10^{-4}$
Skew Normal	MOM-100	$7.6214 \cdot 10^{-4}$	$0.0968 \cdot 10^{2}$	$0.3669 \cdot 10^{-2}$	$1.2272 \cdot 10^{-4}$
Skew Normal	MOM-1000	$7.6214 \cdot 10^{-4}$	$0.0745 \cdot 10^{2}$	$0.1417 \cdot 10^{-2}$	$1.3130 \cdot 10^{-4}$
Skew Normal	MLE	$7.6361 \cdot 10^{-4}$	$0.0544 \cdot 10^{2}$	$0.0271 \cdot 10^{-2}$	$0.0072 \cdot 10^{-4}$
Student's t	MOM-10	$7.5740 \cdot 10^{-4}$	$0.3686 \cdot 10^{2}$	$1.0908 \cdot 10^{-2}$	$0.6775 \cdot 10^{-4}$
Student's t	MOM-100	$7.5740 \cdot 10^{-4}$	$0.0983 \cdot 10^{2}$	$0.3693 \cdot 10^{-2}$	$1.2363 \cdot 10^{-4}$
Student's t	MOM-1000	$7.5740 \cdot 10^{-4}$	$0.0765 \cdot 10^{2}$	$0.1446 \cdot 10^{-2}$	$1.3025 \cdot 10^{-4}$
Student's t	MLE	$7.5856 \cdot 10^{-4}$	$0.0666 \cdot 10^{2}$	$0.0969 \cdot 10^{-2}$	$0.0354 \cdot 10^{-4}$
Uniform	MOM-10	$7.5063 \cdot 10^{-4}$	$0.3712 \cdot 10^{2}$	$1.0959 \cdot 10^{-2}$	$0.6800 \cdot 10^{-4}$
Uniform	MOM-100	$7.5063 \cdot 10^{-4}$	$0.0979 \cdot 10^{2}$	$0.3624 \cdot 10^{-2}$	$1.2153 \cdot 10^{-4}$
Uniform	MOM-1000	$7.5063 \cdot 10^{-4}$	$0.0747 \cdot 10^{2}$	$0.1416 \cdot 10^{-2}$	$1.3044 \cdot 10^{-4}$
Uniform	MLE	$7.5153 \cdot 10^{-4}$	$0.0550 \cdot 10^{2}$	$0.0244 \cdot 10^{-2}$	$0.0055 \cdot 10^{-4}$
Rademacher	MOM-10	$7.5423 \cdot 10^{-4}$	$0.3711 \cdot 10^{2}$	$1.1008 \cdot 10^{-2}$	$0.6820 \cdot 10^{-4}$
Rademacher	MOM-100	$7.5423 \cdot 10^{-4}$	$0.0974 \cdot 10^{2}$	$0.3655 \cdot 10^{-2}$	$1.2231 \cdot 10^{-4}$
Rademacher	MOM-1000	$7.5423 \cdot 10^{-4}$	$0.0747 \cdot 10^{2}$	$0.1427 \cdot 10^{-2}$	$1.3067 \cdot 10^{-4}$
Rademacher	MLE	$7.5527 \cdot 10^{-4}$	$0.0536 \cdot 10^{2}$	$0.0223 \cdot 10^{-2}$	$0.0046 \cdot 10^{-4}$
Skellam	MOM-10	$7.6185 \cdot 10^{-4}$	$0.3726 \cdot 10^{2}$	$1.1077 \cdot 10^{-2}$	$0.6801 \cdot 10^{-4}$
Skellam	MOM-100	$7.6185 \cdot 10^{-4}$	$0.0979 \cdot 10^{2}$	$0.3737 \cdot 10^{-2}$	$1.2379 \cdot 10^{-4}$
Skellam	MOM-1000	$7.6185 \cdot 10^{-4}$	$0.0747 \cdot 10^{2}$	$0.1427 \cdot 10^{-2}$	$1.3141 \cdot 10^{-4}$
Skellam	MLE	$7.6323 \cdot 10^{-4}$	$0.0547 \cdot 10^{2}$	$0.0276 \cdot 10^{-2}$	$0.0074 \cdot 10^{-4}$

Table 1 Mean absolute errors of parameters estimated by various methods from the simulated Ornstein–Uhlenbeck process with true parameters $\mu = 1$, $\tau = 10$ and $\sigma^2 = 10^{-4}$ contaminated by the market microstructure noise with various distributions and variance $\omega^2 = 10^{-8}$.

5 Conclusion

We propose the method of moments estimator of the Ornstein–Uhlenbeck process contaminated by the market microstructure noise for irregularly spaced observations. We compare the proposed estimator with the maximum likelihood estimator in the simulation study. We find that the maximum likelihood estimator is quite robust to misspecification of the distribution of the market microstructure noise. Both methods have their strengths in estimation of specific parameters. For the long-term mean parameter, the method of moments has lower mean absolute error. For the the speed of reversion, instantaneous volatility and noise variance parameters, the maximum likelihood method yields better results. Overall, the method of moments is a decent alternative to the maximum likelihood method for estimation of the parameters of the Ornstein–Uhlenbeck process when the underlying distribution is not specified. The proposed methodology finds its use in modeling financial high-frequency time series exhibiting mean reversion.

Acknowledgements

The work on this paper was supported by the grant No. F4/21/2018 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Aït-Sahalia, Y. & Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton: Princeton University Press. https://doi.org/10.1515/9781400850327

[2] Ball, C. A. & Roma, A. (1994). Target Zone Modelling and Estimation for European Monetary System Exchange Rates. *Journal of Empirical Finance*, **1**(3–4), 385–420. https://doi.org/10.1016/0927-5398(94)90010-8

[3] Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics. *Journal of the Royal Statistical Society: Series B (Methodological)*, **63**(2), 167–241. https://doi.org/10.2307/2680596

[4] Elliott, R. J., Van Der Hoek, J. & Malcolm, W. P. (2005). Pairs Trading. *Quantitative Finance*, **5**(3), 271–276. https://doi.org/10.1080/14697680500149370

[5] Engle, R. F. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica*, **68**(1), 1–22. https://doi.org/10.1111/1468-0262.00091

[6] Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **222**(594–604), 309–368. https://doi.org/10.1098/rsta.1922.0009

[7] Hansen, P. R. & Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, **24**(2), 127–161. https://doi.org/10.1198/073500106000000071

[8] Hautsch, N. (2011). *Econometrics of Financial High-Frequency Data*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-21925-2

[9] Holý, V. (2019). *Financial High-Frequency Data*. University of Economics, Prague. Doctoral thesis. https://insis.vse.cz/zp/portal_zp.pl?podrobnosti_zp=68199

[10] Holý, V. & Tomanová, P. (2018). Estimation of Ornstein-Uhlenbeck Process Using Ultra-High-Frequency Data with Application to Intraday Pairs Trading Strategy. Working paper. https://arxiv.org/abs/1811.09312

[11] O'Hagan, A. & Leonard, T. (1976). Estimation Subject to Uncertainty About Parameter Constraints Miscellanea. *Biometrika*, **63**(1), 201–203. https://doi.org/10.1093/biomet/63.1.201

[12] Schwartz, E. S. (1997). The Stochastic Behaviour of Commodity Prices: Implication for Valuation and Hedging. *The Journal of Finance*, **52**(3), 923–973. https://doi.org/10.1111/j.1540-6261.1997.tb02721.x

[13] Skellam, J. G. (1946). The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations. *Journal of the Royal Statistical Society*, **109**(3), 296. https://doi.org/10.2307/2981372

[14] Uhlenbeck, G. E. & Ornstein, L. S. (1930). On the Theory of the Brownian Motion. *Physical Review I*, **36**(5), 823–841. https://doi.org/10.1103/physrev.36.823

[15] Vasicek, O. (1977). An Equilibrium Characterisation of the Term Structure. *Journal of Financial Economics*, **5**(2), 177–188. https://doi.org/10.1016/0304-405x(77)90016-2

Influence of Membership in the Moravian Aerospace Cluster on the Financial Performance of its Members: Malmquist Index Approach

Natalie Pelloneová[1], Eva Štichhauerová[2]

Abstract. This contribution deals with the influence of membership of a business entity in a cluster organisation on its financial performance. The aim is to verify the hypothesis that the membership in a cluster organisation is connected with increasing financial efficiency of its members in time. In the research sample are included nine charter members of the Moravian Aerospace Cluster. The data are collected for years 2009-2016. For the assessment of financial performance the data envelopment analysis method is applied with two inputs and one output. For inputs were chosen the assets and long-term capital, the economic value added (EVA) was used as the output. Since the EVA achieves both positive and negative values, the variant of radial measure was used. For each company from the sample were calculated the values of the Malmquist index. With help of MI it was possible to quantify the total factor productivity change and to decompose it to technological change and technical efficiency change. The development of the indicators was monitored in time. In conclusion, the results of the research are discussed.

Keywords: cluster, data envelopment analysis, economic value added, financial performance, Malmquist index, Moravian Aerospace Cluster.

JEL Classification: C61, L25, L62
AMS Classification: 90C90, 90B90

1 Introduction

As part of the co-operation among several completely independent enterprises, broad business networks or clusters may form [6]. The idea of clustering companies is based on Alfred Marshall's concept of industrial districts from the end of the 19th century [1]. However, the development of clusters did not begin until the early 1990s, when Michael Eugene Porter's ground-breaking book entitled "The Competitive Advantage of Nations" was published [3]. In his contribution, Porter [14] coined the term cluster and described it as a new way to boost the competitiveness of companies, innovation, and industrial and economic development, as well as a tool to improve the performance of individual participating companies. Porter [14] defines a cluster as "a geographically proximate group of interconnected companies, specialised suppliers, service providers and associated institutions in a particular field that compete and co-operate with each other and are linked by commonalities and complementarities". In terms of structure, cluster members most often include manufacturing and trading companies, knowledge institutions and support institutions.

The main economic effect of the operation of clusters is their positive impact on their member companies' competitiveness and performance [12]. Clusters allow participating companies to improve their performance through specialised suppliers, technologies, and information. Cluster performance can be managed and measured in different ways – according to [12], the most commonly used methods for measuring cluster performance include: the Cluster Initiative Performance Model, the Norwegian Cluster Benchmarking Model, the Balanced Scorecard method, and Economic Value Added. However, it is safe to conclude that the number of available technical publications dealing with the impact of membership in cluster organisations on companies' financial performance is currently limited.

The present paper examines the effect of business entities' membership in a selected cluster organisation on their financial performance. The research builds on a previous paper [13] that compared the performance of members of the Moravian aerospace cluster with non-members over the 2010–2016 period in order to determine whether member companies of the Moravian aerospace cluster achieved better financial performance than non-

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz
[2] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, eva.stichhauerova@tul.cz

member companies. In subsequent research, several fundamental changes have been made. The time series was extended to include an additional period (2009), and only the effect of this cluster on the financial performance of member business entities was examined. The objective was to evaluate the effect of this cluster organisation on it member business entities' financial performance, and to use the Malmquist Index to confirm the assumption that business entities' membership in a cluster organisation translates into improved financial performance in a time series. Economic Value Added (EVA) was chosen as the measure of performance.

2 Malmquist index

Basic data envelopment analysis (DEA) models do not take into account trends or changes in the efficiency of entities' activities over time. This deficiency is eliminated using the Malmquist Index (MI), which evaluates changes in efficiency over time. The MI is based on DEA models and it is an important indicator that is used to measure changes in the relative efficiencies of DMUs in different time periods. This index was first introduced by Malmquist in 1953 within the context of consumer theory. In 1982, Caves expanded the index to measure productivity changes in two time periods as the distance between the decision-making unit (DMU) and the frontier for each period [4]. Fare et al. [7] further developed the MI in the context of performance assessment as an empirical index. The MI can be decomposed into two parts. The first component measures technical efficiency changes (E) and the second component measures technological changes (T) between periods t and $t + 1$.

In this paper, input-oriented MI was used, which can be expressed by equation (1). Where x^t are inputs in period t, y^t are outputs in period t, x^{t+1} are inputs in period $t + 1$, y^{t+1} are outputs in period $t + 1$, E_q is the change in the relative efficiency of unit q with respect to other units between periods t and $t + 1$, T_q is the change in production possibility frontier as a result of technology development between periods t and $t + 1$, or in other words technological change. In this equation, the MI is calculated as the product of E_q and T_q, and it provides several advantages over other indices.

$$MI_q(x^{t+1}, y^{t+1}, x^t, y^t) = E_q T_q \tag{1}$$

Components E_q and T_q are given by equations (2) and (3). Where $D_q^t(x^t, y^t)$ is production unit U_q efficiency given the existing technology in period t with inputs and outputs from period t, $D_q^t(x^{t+1}, y^{t+1})$ is production unit U_q efficiency given the existing technology in period t with inputs and outputs from period $t + 1$, $D_q^{t+1}(x^t, y^t)$ is production unit U_q efficiency given the existing technology in period $t + 1$ with inputs and outputs from period t, $D_q^{t+1}(x^{t+1}, y^{t+1})$ is production unit U_q efficiency given the existing technology in period $t + 1$ with inputs and outputs from period $t + 1$. The values $D_q^t(x^t, y^t)$, $D_q^t(x^{t+1}, y^{t+1})$, $D_q^{t+1}(x^t, y^t)$, $D_q^{t+1}(x^{t+1}, y^{t+1})$ can be obtained by solving DEA models, either for the assumption of constant or variable returns of scale (VRS).

$$E_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{2}$$

$$T_q = \sqrt{\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}} \tag{3}$$

A value of $MI_q > 1$ indicates an increase in productivity; $MI_q = 1$ means there has been no productivity change; and $MI_q < 1$ means a decrease in productivity. If the value of $E_q > 1$, it means that the company has improved its technical efficiency. If it equals one, there has been no change in technical efficiency, and if $E_q < 1$, then the company's technical efficiency has deteriorated, and the company has moved away from the efficiency frontier. T_q values can be interpreted in a similar way, if $T_q > 1$, there has been a positive shift or technological advancement; if it equals one, there has been no technology change within the company; a value of $T_q < 1$ indicates a negative shift or technological decline.

DEA was used to calculate the values of the distance functions. In standard DEA models, it is assumed that all inputs and outputs are non-negative. The output used, i.e. EVA, can take both positive and negative values, which is contrary to the requirements of traditional DEA models. In this paper, the variant of radial measure (VRM) proposed by [5] was used for eliminating the problem of non-positive values, because the VRM model is units invariant and can deal with every case of negative values presence in the dataset under assessment. This model was formulated for the condition of VRS [5]. An input-oriented VRM model operating under VRS is shown below in (4) to (8), where λ_j are weights of all DMUs. The constraint (7) is the condition of convexity which is kept under VRS. The modification to the model consists in using the absolute values of inputs (outputs)

instead of their actual values. It can be noticed that β measures how much an observed DMU should improve in order to reach the efficient frontier, in other words it represents the inefficiency measure.

$$Max\ \beta \tag{4}$$

subject to:

$$X\lambda + \beta|x_0| \leq x_0 \tag{5}$$
$$Y\lambda \geq y_0 \tag{6}$$
$$e^T\lambda = 1 \tag{7}$$
$$\lambda \geq 0 \tag{8}$$

3 Data and Methodology

The cluster chosen for this case study was the Moravian Aerospace Cluster, which was set up in March 2010 under the initiative of CzechInvest and it has the legal form of an association. Based in Kunovice, the cluster operates in southern and south-eastern Moravia, where the aerospace industry has a long tradition. The cluster brings together legal entities operating in the field of aircraft manufacturing, related equipment and related services, as well as research organisations engaged in research and development in the field of technical sciences [10]. The basic data source was information obtained from the financial statements (balance sheet, profit and loss statement) of the Moravian Aerospace Cluster's member organisations for the 2009–2016 period, which were obtained from the MagnusWeb database [2]. The period under review was selected taking into account the development of the cluster and the beginning of its operation, while allowing for the fact that the effects of cluster membership can be expected with a certain delay. Most companies have not yet published their financial results in the Commercial Register, which is why the time series ends in 2016.

The research was carried out in the following steps:

1. Compiling a list of the companies to be evaluated - as the first step, it was necessary to create a database of the member entities of the cluster organisation selected. In the period analysed, the Moravian Aerospace Cluster had 25 members in total. Since the research focused on evaluating financial performance, only business entities were included as they are generally expected to seek such goals as maximising their profit or value added. In the period analysed, the cluster had 18 businesses-entity members with conducting heterogeneous activities. In the research in question, it was only possible to compare companies that had the same or similar line of business, form the core of the cluster and which had been members of the cluster for the same period of time – these are the only companies that could be considered the homogeneous core of the cluster. In order to evaluate the companies' financial performance in a time series, it was also necessary to define the homogeneous core of the cluster, i.e. the founding member enterprises. Based on an analysis of the line of business in accordance with the NACE classification, the core of the cluster consisted of nine manufacturing companies in CZ-NACE sectors 30300, 25600 and 24500.

2. Gathering financial statements - it was necessary to obtain the required data from the above companies' financial statements (especially from their balance sheets and profit and loss statements) for 2009–2016. The success rate of obtaining the financial statements was 100%. Financial statements were obtained for all 9 companies for all years.

3. Calculating economic value added - for each of the entities, the indicator entitled economic value added (hereinafter EVA) was subsequently calculated in accordance with the Ministry of Industry and Trade methodology [11]. The MIT calculates the EVA indicator using an equity-based approach where EVA is defined as the product of equity E and 'spread' (return on equity ROE minus alternative cost of equity r_e), see relation (9). EVA can take both positive and negative values. A positive EVA means that the company has generated value for its owners. If EVA is negative, the value of the company decreases. This indicator should provide a basic picture of the financial performance of companies within the Moravian Aerospace Cluster.

$$EVA = (ROE - r_e) \cdot Equity \tag{9}$$

4. Input and output specification - total assets and long-term capital were selected as inputs for the model. Long-term capital is the sum of the following balance sheet items: equity, long-term bonds issued and long-term bank loans. The output is economic value added (EVA).

5. Determination of technical efficiency values - for each enterprise within the set, an efficiency score D_q was calculated, assuming VRS. For each enterprise identified as efficient, the super-efficiency score was calculated.

6. Calculation of the Malmquist Index - for each DMU and each period, the values of the distance functions and of each component of the Malmquist Index were determined in the MaxDEA 7 Ultra software environment, using equations (2) and (3). Finally, the value of the Malmquist Index was calculated using equation (1).

4 Research results and discussion

Table 1 shows the trends in the technical efficiency score of the cluster's founding members for the years 2009 to 2016 and, where relevant, the super-efficiency score of the enterprises that were identified as efficient (a value greater than or equal to 1). For this purpose, Andersen and Petersen super-efficiency model was applied. In almost every year, three efficient units were identified within the set. It is obvious that VR GROUP, a.s. was the only company that was identified as an efficient unit in each year within the 2009–2016 period. It is a subsidiary of the state-owned enterprise LOM PRAHA s.p., which was established by the Ministry of Defence of the Czech Republic. Its line of business is the design and implementation of comprehensive training and education systems for the aerospace industry and others. The value of technical efficiency increased by about 20% per year until 2014, even showing an increase as high as 71% in 2013/2014. However, it then continually decreased until the end of the period under review (by 14% for 2015/2016). Most recently, the company was publicly discussed in connection with a proposed contract for a new live simulation system for the Czech Army, which was yet again awarded to the Swedish company Saab without a public tender. Since the Czech government has identified supporting the domestic arms industry as one of its priorities, many consider this decision to be incorrect [8].

	2009	2010	2011	2012	2013	2014	2015	2016
5M s.r.o.	0.53	0.68	0.12	0.11	0.17	0.70	0.26	0.86
Aircraft Industries, a.s.	1.00	1.00	0.02	0.02	0.02	0.02	0.02	0.02
ALUCAST, s.r.o.	1.20	0.64	0.73	0.46	0.89	0.89	0.44	0.71
C.S.O., spol. s r.o.	0.75	1.88	1.43	1.08	1.31	2.06	0.71	1.08
EVEKTOR, spol. s r.o.	0.13	0.11	1.00	1.00	2.32	0.08	0.12	0.05
KOVOVÝROBA HOFFMANN, s.r.o.	0.05	0.04	0.04	0.03	0.05	0.07	0.83	1.00
MESIT holding, a.s.	0.02	0.02	0.02	0.02	0.03	0.04	0.02	0.04
První brněnská strojírna Velká Bíteš, a. s.	0.03	0.03	0.02	0.02	1.00	1.00	1.00	0.51
VR GROUP, a.s.	3.78	4.72	5.97	6.70	8.17	14.04	6.46	5.55

Table 1 Efficiency and super-efficiency score

For the following enterprises, significant changes in trends were identified: C.S.O., spol. s r. o.; Aircraft Industries, a.s.; První brněnská strojírna Velká Bíteš, a.s.; EVEKTOR, spol. s r.o.

C.S.O., spol. s r.o., was identified as an efficient unit in almost every year (except for 2015). The company keeps investing into expanding its production capacities, and the construction of production and storage facilities was completed in 2015. Therefore, while technical efficiency decreased by 65% in 2014/2015, in terms of technology, there was an increase of 56%. In the subsequent 2015/2016 period, technical efficiency increased by 52%.

Since 2008, Aircraft Industries, a.s. has been majority-owned by the Russian Ural Mining and Metallurgical Company, with most of its exports going to the Russian Federation. While it was still identified as an effective unit in 2009 and 2010, both efficiency and technical efficiency decreased by approximately 98% in 2010/2011. In 2013, the technical efficiency of this company started to increase again (by 6.5% in 2012/2013, 22% in 2013/2014 and 15% in 2014/2015). At the end of the period under review (2016), there was yet again a year-on-year decline of ca 6%. In 2016, the company was in danger of going bankrupt, but the insolvency petition was rejected, and the company agreed with its creditors on a repayment schedule. It can be assumed that the company will continue to grow in the coming years because, since 2015, aircraft have regained their position as an important item of Czech export to the Russian Federation, and the civil aerospace industry is currently also considered a very promising sector for Czech exports to Russia in the long term.

První brněnská strojírna Velká Bíteš, a.s. specialises in the manufacture of high-speed turbine machines and it is also a leading European precision casting foundry. Along with many other companies, it was hit by the economic crisis in 2009, which also translated into a very low efficiency score (ca 0.03 in 2009 and 2010). In 2012/2013, the company reported a record year-on-year increase in technical efficiency – about 47-fold, and the MI was 33.8. In these years, the company received several prestigious awards – it won the 2012 and 2013 Company of the Year competition in the Vysočina Region, and even at the national level in 2012. In 2017, the company was recognised as the Responsible Company of the Year, among other things for offering equal pay for

men and women, providing a high-quality range of benefits, supporting employee volunteering and actively seeking to eliminate negative environmental impacts. The company's activities are well strategically targeted, aiming for long-term support and development of the local community and region.

Significant changes in trends were also identified in the case of EVEKTOR, spol. s r.o., a company that develops and builds small aircraft. In the 2010/2011 period, there was a nine-fold increase both in terms of overall efficiency (an MI of 9.02) and, above all, in terms of technical efficiency (an E of 9.30). The company thus reached the production-possibility frontier and remained there until 2013. Even in 2012/2013, its technical efficiency still increased by 131%. However, in the subsequent period of 2013/2014, its efficiency declined: the company fell well below the production-possibility frontier (an efficiency score of 0.076) and both its efficiency and technical efficiency decreased by about 97%. The company remained an inefficient unit until the end of the period under review, i.e. 2016. The deterioration may be due to the situation surrounding the unsuccessful implementation of the new EV-55 aircraft project, for which the Malaysian company Aspirasi Pertiwi promised an investment in 2014, while also acquiring an 18% stake and decision-making rights in EVEKTOR, spol. s r.o. However, no investment was made and as early as 2015, the investor already owed CZK 220 million. It has been speculated that the Malaysian investors were in fact interested in moving away from aircraft manufacturing and focusing on the automotive industry. However, one cannot avoid noticing that when the investor's main representative and former Malaysian Prime Minister Mohammad Mahathir resigned his position at the Proton automobile manufacturing company, the Malaysians stopped showing any interest in the Czech company [15]. Anyway, EVEKTOR, spol. s r.o. is currently facing financial difficulties.

The cluster's average efficiency score (D) for each year is shown in Table 2 on the left (the arithmetic mean without a super-efficiency model was used). At first sight, the fluctuating values do not show any clear trend.

Year	D	Period	MI	E	T
2009	0.5004	2009/2010	1.20	1.05	1.14
2010	0.5013	2010/2011	0.71	0.66	1.09
2011	0.4390	2011/2012	0.85	0.87	0.97
2012	0.4060	2012/2013	1.52	2.18	0.69
2013	0.5720	2013/2014	0.85	0.99	0.87
2014	0.5331	2014/2015	1.16	0.91	1.28
2015	0.4896	2015/2016	0.78	1,13	0.69
2016	0.5768	**Mean**	**0.98**	**1.04**	**0.94**

Table 2 Average efficiency score and Malmquist index

Table 2 also provides an overview of trends in the average values for MI and its components for the cluster in the 2009 to 2016 period. Given that the MI is constructed as a multiplicative model, the geometric mean was used to calculate these average values. At the beginning of the period, i.e. in 2009/2010, the MI value was 1.20, which indicates a 20% increase in the average efficiency of the cluster members. By analysing the components of the MI, it was found that technological changes accounted for most of the increase (14%), while the technical efficiency of cluster members increased by ca 5%. The biggest year-on-year increase in MI (an increase of 52%) took place in 2012/2013. The index decomposition then shows that, in the same period, the technical efficiency of cluster members increased by 118% on average. The most recent major changes in efficiency occurred in 2015/2016 – the MI value decreased by 22%, but this was due to a negative technology change (a 31% decline), while technical efficiency increased by 13% on average. While the MI value decreased on average by 2% over the entire 2009–2016 period, the decomposition of the index shows that the cluster's technical efficiency increased by 4% on average, i.e. it improved.

5 Conclusion

This paper dealt with evaluating the trends in the financial performance of nine founding members of the Moravian Aerospace Cluster in the 2009–2016 reference period. The paper used DEA analysis, namely the input-oriented BCC model expanded to include the Malmquist Index, as a tool to compare the trends in efficiency over time. Total assets and long-term capital were selected as inputs, the EVA indicator (calculated in accordance with the methodology of the Ministry of Industry and Trade of the Czech Republic) was selected as the output. In almost every period, three businesses were identified as efficient within the cluster. First, an individual analysis of the trends in the efficiency of selected cluster members was performed and possible causes of significant changes in trends were discussed.

Even though the cluster's average efficiency decreased by 2 per cent over the entire 2009–2016 period, its technical efficiency – i.e. the ability to efficiently transform inputs into outputs – improved by 4 per cent. This supports the claim that business entities' membership in a cluster organisation translates into their improved financial performance in a time series.

While the reference period for this study ends in 2016, the authors of the paper will continue their research in subsequent years. It can be assumed that the efficiency of Czech enterprises in the aerospace industry might increase in connection with export possibilities in the future. The latest 2018/2019 edition of the Global Industry Opportunities Map 2018/2019 lists 32 countries around the world that represent export opportunities within the civil aerospace industry [9]. The countries with the biggest export potential are the US and France (collectively ca USD 1 654.3 trillion), followed by Canada (ca USD 221 trillion), Russia (ca USD 79.3 trillion) and Malaysia (ca USD 64.4 trillion). The export items that account for the largest share include radar components and other aircraft such as helicopters and ultralight aircraft. Historically, Czech aerospace products have had excellent reputation and tradition, which could help Czech enterprises to take advantage of these opportunities.

Acknowledgements

Supported by the grant No. GA18-01144S „An empirical study of the existence of clusters and their effect on the performance of member enterprises" of the Czech Science Foundation.

References

[1] Belussi, F. & Caldari, K. (2008). At the Origin of the Industrial District: Alfred Marshall and the Cambridge School. *Cambridge Journal of Economics*, 33(2), 1–30.
[2] Bisnode. (2018). *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Praha: Bisnode Česká republika, 2018 [cit. 2018-06-25]. Dostupné z: https://magnusweb.bisnode.cz
[3] Boja, C. (2011). Clusters Models, Factors and Characteristics. *International Journal of Economic Practices and Theories*, 1(1), 34–43.
[4] Caves, D. W., L. R. Christensen & W. E. Diewert. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6), 1393–1414.
[5] Cheng, G., Zervopoulos, P. & Qian, Z. (2011). A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. MPRA Paper 30951, University Library of Munich, Germany.
[6] Dvořáček, J. & Tyll, L. (2010). *Outsourcing a offshoring podnikatelských činností*. Praha: C.H. Beck.
[7] Fare, R., et al. (1994). Productivity developments in swedish hospitals: a malmquist output index approach. In: A. Charnes, W. W. Cooper, A. Y. Lewin & L. M. Seiford (Eds.), *Data envelopment analysis: theory, methodology and applications* (pp. 253–272). Boston: Kluwer Academic Publishers.
[8] Hrubeš, K. (2018). Vojáci potřebují nový systém živé simulace. Tendr bez soutěže budí pochyby. [online]. Praha: MAFRA, a. s., 2019 [cit. 2019-04-01]. Available from: https://www.idnes.cz/zpravy/domaci/tendr-verejna-zakazka-armada-system-zive-simulace-namestek-ministr-obrany-lubomir-metnar.A180925_162423_domaci_ane.
[9] Mapa globálních oborových příležitostí 2018/2019. (2018). Civilní letecký průmysl. [online]. Praha: Ministerstvo zahraničních věcí ČR, 2019 [cit. 2019-04-01]. Available from: https://www.mzv.cz/ekonomika/cz/udalosti/mapa_globalnich_oborovych_prilezitosti_1.html
[10] Moravian Aerospace Cluster. (2018). *O klastru.* [online]. Kunovice: Moravian aerospace Cluster, 2018 [cit. 2018-10-01]. Available from: http://www.aero-cluster.cz/?q=cs/.
[11] MPO. (2018). Finanční analýza podnikové sféry za rok 2015. Ministry of Industry and Trade, Prague, 2018.
[12] Pavelková, D., et al. (2009). *Klastry a jejich vliv na výkonnost firem*. Praha: Grada Publishing.
[13] Pelloneová, N & Štichhauerová, E. (2018). Vliv členství v klastru na finanční výkonnost podniků: Případová studie z leteckého průmyslu. In Magnanimitas (Eds.), *Proceedings of the International Masaryk Conference for Ph.D. Students and Young Researchers 2018* (pp. 86–94). Hradec Králové: Magnanimitas.
[14] Porter, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, 76(6), 77–90.
[15] Tomeš, M. (2018). Projekt českého letadla je před krachem. Investice od malajsijského expremiéra nedopluly. [online]. Praha: CZECH NEWS CENTER a.s., 2019 [cit. 2019-04-01]. Available from: https://www.e15.cz/byznys/prumysl-a-energetika/projekt-ceskeho-letadla-je-pred-krachem-investice-od-malajsijskeho-expremiera-nedopluly-1346306.

An Assessment of Regional Variations: An Application to Polish Regions

Martin Dlouhý[1], Jakub Hanousek[2]

Abstract. To reduce regional variations in health resources is major goal of the national health policy. The variations in regional health resource capacities are observed in many European health systems that are publicly financed and highly regulated. The objective of this paper is to show three methods of an assessment of regional health resource capacities: the separate evaluation, the common weights model, and the production frontier model. The presented methods are applied to the regional distribution of doctors and nurses in Poland in 2015. The data come from the Eurostat regional statistics database that divides Poland to 16 NUTS 2 regions. The common weights and production frontier models that take into account the possibility of resource substitution give on average higher capacity scores and show lower differences between regional capacities.

Keywords: health system, inequality, resource allocation, common weights model, data envelopment analysis, production frontier model.

JEL Classification: I14, D63
AMS Classification: 91B32

1 Introduction

The subject of health economics is the allocation of scarce health resources to and within the health system. The objective of the resource allocation is the optimal distribution and combination of these scarce health resources to achieve the maximum desired output. In the health system, we distinguish two criteria of the resource allocation: efficiency and equity. Efficiency is concerned with whether resources are allocated to the right activity to achieve maximum benefits without any waste. On the other hand, equity is a little bit complicated term that relates to the fairness of access to resources. In this paper, we deal with only one dimension of equity—a fair (equal) geographical distribution of health resources. Because, the goals of efficiency and equity can be in conflict, difficult decisions have to be made at some situations.

Resource allocation within the health system can be analysed according to several criteria: (a) the source of financing, (b) the type of service, (c) the type of provider, (d) the disease category, and (e) the geographical area. In this study, we will focus on the geographical (regional) variations. To reduce unexplained regional variations in access to health resources, in health services utilisation, and in health expenditures are major goals of the national health policy (see for example [2, 3, 7, 9]). In fact, this goal is an essential element of publicly financed health system that affects the organization of the whole health system. Since the majority health systems of the world is publicly funded (taxes or public health insurance) and highly regulated by the government, health insurance funds and other regulatory institutions, the observed regional variations clearly represent a failure of national health policy.

In this study, we will focus on measurement of variations in regional health resource capacities. It may be a surprise that the variations in regional health resource capacities are observed in many European health systems in spite of the fact that these systems are publicly financed and regulated. An analysis of the regional distribution of health resources is one of the tools of evaluation of equal access to health care. The usual analytical approach to an assessment of regional differences is to evaluate each health resource separately. The objective of this study is to show alternative methods of regional comparisons to the method of separate evaluation.

The presented methods are illustratively applied to the regional distribution of health resources in Poland. Although the both authors of this paper come from the Czech Republic, the authors decided to use the Polish data because of higher number of regions, which can potentially better serve for observing the behaviour of methods.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, 4 Winston Churchill Sq., Prague, Czech Republic, e-mail: dlouhy@vse.cz
[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, 4 Winston Churchill Sq., Prague, Czech Republic, e-mail: xhanj52@vse.cz.

2 Methods

2.1 Separate Evaluation

The most common method of assessment is to evaluate regional resource capacities for each health resource separately. The researchers usually compare regional capacities of doctors, nurses, hospital beds and the numbers of medical technology. No substitution among health resources is considered and therefore the total sum of health resource capacities does not make sense. The separate evaluation of each health resource is simple and easily understandable by all health policy stakeholders.

However, the separate evaluation has some disadvantages, especially in ignoring possible substitutions among health resources. For a simplicity, let us assume that more is better, although it is not always the case. Hence the population in region *A* has surely better access to health care than the population in region *B* only in the case if region *A* mathematically dominates region *B*, i.e. capacities of all health resources in region *A* are higher than the capacities in region *B*. In the case that some capacities of health resources are higher in region *A* and some capacities are higher in region *B*, the separate evaluation method is not able to decide which regional population is better-off.

2.2 Common Weights Model

Unlike the method of separate evaluation (section 2.1), we assume that health resources are, at least to some extent, substitutes, so the total weighted sum of health resource capacities can have a health policy interpretation. It is assumed that the lower health resource capacities in one resource category can be compensated by higher capacities in another resource category. In such a case, the size of geographical inequality is expected to be lower than the inequality calculated by the separate evaluation.

To cope with cases with multiple health resources, the total health resource capacity is obtained by the weighted sum of individual health resource capacities. A comparison of health resource capacities in this situation is a problem of multiple criteria decision making, specifically of the weighted sum model in which the key issue is setting of relative resource weights. Such resource weights are then used nationwide for all regions. The question is how to obtain values of these resource weights. A survey among experts that are able to give us their subjective views is one possibility. The alternative possibility is to calculate weights "objectively" by the mathematical model.

The objective function of the presented model is a maximization of the total weighted sum of health resource capacities. This model is known as the common weights model. Let us have a country with *n* regions that use *m* inputs to serve regional population. In this case we assume that regional population is the single output that can be omitted from the model if the inputs are expressed in the numbers per inhabitant. The mathematical formulation of the common weights model is:

$$
\begin{aligned}
&Maximize \quad \sum_{j=1}^{n} \varphi_j \\
&subject\ to \quad \varphi_j = \sum_{i=1}^{m} v_i x_{ij} \qquad j = 1, 2, \ldots, n, \\
&\qquad \sum_{i=1}^{m} v_i x_{ij} \leq 1, \qquad j = 1, 2, \ldots, n, \\
&\qquad v_i \geq \varepsilon, \quad i = 1, 2, \ldots, m,
\end{aligned}
\tag{1}
$$

where φ_j is the normalized capacity score of the region j (the score is set to be 1 or 100 for the region with the highest total capacities and is lower for the others), x_{ij} is the capacity of health resource i per 1000 inhabitants in region j, ε represents an infinitesimal constant that assures that the weight for each health resource is greater than zero. The resource weights v_j are variables in the model.

The idea of the common weights model is that the national health system as a whole tends to optimize the value of total health resource capacity and that relative ratios of health resource capacities are the same for all regions in the given country.

2.3 Production Frontier Model

More flexible approaches to estimate resource weights and the rate of substitution can be based on the production frontier [4]. In this method, the weights of health resources differ for each region. In the production frontier model, the health resources can be considered as inputs and the served population (as a basic measure of health need) is the single output. In this case, the production function is estimated by the "resource efficient" units with minimal amounts of resources. So the capacities of "resource inefficient units" are compared to the most efficient production frontier. In the second form of the production frontier, the regional population is modelled as the single input and health resources are model outputs. In this alternative that we will use in the paper, the production frontier is estimated by the most "resource inefficient" units, i.e. the units with the maximal amount of resources.

The production frontier can be estimated by the deterministic frontier analysis or stochastic frontier analysis, which are econometric methods [8], or by data envelopment analysis, which is method based on the theory of mathematical programming (see, for example, publications [5, 6]). In this paper, we will use the data envelopment analysis for estimation (calculation) of the production frontier.

Data envelopment analysis (DEA) was developed by Charnes, Cooper and Rhodes in 1978 [1] and constructs the production frontier and evaluates the technical efficiency of production units. DEA is based on the theory of mathematical programming and estimates the production frontier as the piecewise linear envelopment of the data. The production unit uses a number of inputs to produce outputs. The technical efficiency of the production 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.

The DEA model permits each production unit to choose its input and output weights to maximize its technical efficiency score. A technically efficient production unit is able to find such weights that the production unit lies on the production frontier. The production frontier represents the maximum amounts of output that is produced by given amounts of input (the output maximization DEA model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (the input minimization DEA model).

Let us have n production units that use m inputs to produce r outputs. The formulation of the input-oriented version of the constant returns-to-scale DEA model for production unit q is:

$$
\begin{aligned}
&Maximize \quad \varphi_q = \sum_{k=1}^{r} \varphi_k y_{kq} \\
&subject\ to \quad \sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, 2, \ldots., n, \\
&\qquad \sum_{i=1}^{m} v_i x_{iq} = 1, \\
&\qquad u_k \ge \varepsilon, \quad k = 1, 2, \ldots, r, \\
&\qquad v_i \ge \varepsilon, \quad i = 1, 2, \ldots, m,
\end{aligned}
\tag{2}
$$

where φ_q is the technical efficiency score (the normalized capacity score of the region in our case), x_{ij} is the amount of input i used by unit j, y_{kj} is the amount of output k produced by unit j, and ε represents an infinitesimal constant that assures that the weight for each health resource is greater than zero. The output weights u_k and input weights v_i are variables in the DEA model. In the input-oriented DEA model, the efficiency score φ_q is one if the unit q is technically efficient, and is lower than one if the unit is technically inefficient. The efficiency score calculates a size of input reduction that makes production unit q technically efficient. In the output-oriented DEA model, the efficiency score is one if the unit q is technically efficient, and is greater than one if the unit is technically inefficient.

3 Illustrative Application

Three presented assessment methods that are described in the chapter 2 were applied to the example of Poland. The Polish data come from the Eurostat regional statistics database that divides the territory of Poland to 16 NUTS

2 regions. The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the European Union. The data come from year 2015. The regional capacities of two health resources, doctors and nurses, are evaluated. In 2015, Poland had a population 38 million inhabitants that were served by 88 437 doctors (2.3 doctors per 1000 inhabitants) and by 219 845 nurses (5.8 nurses per 1000 inhabitants). The regional data that were used in our calculations are presented in Table 1.

Region	Doctors	Nurses	Population
Lódzkie	6,676	13,847	2,489,844
Mazowieckie	14,497	31,709	5,309,778
Malopolskie	7,668	20,169	3,324,579
Slaskie	11,077	28,303	4,535,705
Lubelskie	5,319	14,271	2,126,310
Podkarpackie	4,467	13,974	2,084,264
Swietokrzyskie	2,916	8,224	1,247,951
Podlaskie	3,011	7,364	1,162,369
Wielkopolskie	5,231	15,525	3,447,689
Zachodniopomorskie	4,116	8,593	1,688,486
Lubuskie	2,035	5,396	1,007,442
Dolnoslaskie	6,540	16,274	2,868,338
Opolskie	1,946	5,412	956,913
Kujawsko-Pomorskie	4,813	11,935	2,065,846
Warminsko-Mazurskie	2,983	7,550	1,418,541
Pomorskie	5,142	11,299	2,271,559
Poland	88,437	219,845	38,005,614

Table 1 Regional Data, Poland, 2015

In order to better compare the results achieved by different methods, the regional capacity scores were normalized to the maximal regional capacity being 100 (see Table 2). In the case of the production frontier model we used the input-oriented constant returns-to-scale DEA model (2) with the regional population as the single input and the number of doctors and the number of nurses as two outputs. All calculations were carried out in MS Excel with the help of the Solver and DEA add-in programs.

The Polish regions with highest resource capacities are the Mazowieckie and Lubelskie regions. One region has the maximal number of doctors per 1000 inhabitants (Lubelskie) and one region has the maximal number of nurses per 1000 inhabitants (Mazowieckie). On the other hand, the minimal health resource capacity is observed by all methods in the Wielkopolskie region. Not surprisingly, the common weights and production frontier models that take into account the possibility of resource substitution show on average higher scores and thus lower differences between regional health resource capacities (see Table 2). This can be, for example, observed in the cases of the Podlaskie and Dolnoslaskie regions.

Region	Separate Evaluation		Common Weights Model	Production Frontier Model
	Doctors	Nurses		
Lódzkie	98	83	96	98
Mazowieckie	100	89	100	100
Malopolskie	84	90	91	91
Slaskie	89	93	96	96
Lubelskie	92	100	100	100
Podkarpackie	78	100	92	100
Swietokrzyskie	86	98	96	98
Podlaskie	95	94	99	99
Wielkopolskie	56	67	64	67
Zachodniopomorskie	89	76	88	89
Lubuskie	74	80	80	80
Dolnoslaskie	84	85	88	88
Opolskie	74	84	83	84
Kujawsko-Pomorskie	85	86	90	90
Warminsko-Mazurskie	77	79	82	82
Pomorskie	83	74	83	83
Average	84	86	89	90

Table 2 Normalized Resource Capacities

4 Conclusion

We focused on measurement of variations in regional health resource capacities. In the paper, we have presented three alternative methods of assessment of variations in regional health resource capacities: the separate evaluation, the common weights model, and the production frontier model based on the data envelopment analysis. Illustrative calculations were carried out on regional data from Poland that is divided to 16 NUTS 2 regions. It is observed that both the common weights model and production frontier model that take into account the possibility of resource substitution show on average higher capacity scores and thus lower differences between regional health resource capacities. Hence both health services researchers and health policy makers should always consider the possibility of resource substitutions in assessment of regional variations.

A comparison of resource capacities can be seen as a problem of multiple criteria decision making in which the main issue is setting of resource weights. In this paper, the calculation of resource weights were carried out objectively by the common weights and production frontier models. In the future research, we plan a survey among health policy experts and health managers that give us their subjective views. Combining both the expert knowledge and quantitative approaches can bring us new research insights.

Acknowledgements

Martin Dlouhý was supported by the project no. 19-08985S, funded by the Czech Science Foundation, and Jakub Hanousek was supported by the project no. F4/66/2019, funded by the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague

References

[1] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the Inefficiency of Decision Making Units. *European Journal of Operational Research*, 2, 429–444. doi: 10.1016/0377-2217(78)90138-8.

[2] de Looper, M. & Lafortune, G. (2009). Measuring Disparities in Health Status and in Access and Use of Health Care in OECD Countries. *OECD Health Working Papers*. No. 43, Paris: OECD Publishing.

[3] Dlouhý, M. (2005). Health care regulation: impact on the supply of outpatient services in the Czech Republic. *Prague Economic Papers*, 14(3), 267–276.
[4] Dlouhý, M. (2018). Measuring Geographic Inequalities: Dealing with Multiple Health Resources by Data Envelopment Analysis. *Frontiers in Public Health*, 6, article 53. doi: 10.3389/fpubh.2018.00053.
[5] Dlouhý, M., Jablonský, J. & Zýková, P. (2018). *Analýza obalu dat* [Data Envelopment Analysis, in Czech]. Prague: Professional Publishing, 2018. ISBN 978-80-88260-12-7.
[6] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů* [Models of Efficiency Evaluation and Resource Allocation, in Czech]. Prague: Professional Publishing.
[7] Johnston, G. & Wilkinson, D. (2001). Increasingly inequitable distribution of general practitioners in Australia, 1986–96. *Australian and New Zealand Journal of Public Health*, 25(1), 66–70.
[8] Kumbhakar, S. C. & Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge: Cambridge University Press.
[9] OECD. (2014). Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance? *OECD Health Policy Studies*. Paris: OECD Publishing. ISBN 978-92-64-21659-4.

Efficiency and Investment. The Czech Industry Case.

Václava Pánková[1]

Abstract. A sample of 24 representative firms in the Czech economy is a subject of a study for their technical efficiency and subsequently for their willingness to invest. The first concept is accomplished with the help of the frontier production function. The latter one is based on the value of Tobin´s Q, defined as the ratio of the market value of business capital assets to their replacement value; if it is greater than one, Q indicates the profitability of further investment.
The followed firms differ in their technical performance, but all of them are profitable and this might be their motivation to invest. As a consequence, the concept of governmental investment incentives should possibly be reconsidered.

Keywords: frontier production function, Tobin´s Q, panel data, VAR model

JEL Classification: C10, C13, C33, D2
AMS Classification: 90B30, 62F10

1 Introduction

The Czech Republic (CR) is a small open economy with a high demand for investment, part of which comes from abroad. The foreign investors, especially, have no reason to stay if they do not earn profits. Therefore, it seems probable that they produce at an acceptable degree of efficiency, but it is not clear if they intend to invest repeatedly or to explore actual resources only and leave the country. Due to today´s high capital mobility such a production behaviour is a current practice.

In this article, a sample of 24 firms representing the manufacture of motor vehicles, trailers and semi-trailers and other transport equipment in the CR are treated. The chosen firms are parts of international concerns (e.g., Škoda Auto belongs to Volkswagen), represent foreign capital invested in the CR (e.g., Hyundai) or are speculated to be the subject of potential foreign acquisition (Škoda Transportation). In fact, the sample makes up such a substantial part of the Czech industry that the collapse of the sample firms will most probably be an economic disaster for the country. The data comes from the database Amadeus32 (Bureau van Dijk, [1]) and covers the years from 2007 to 2016 (last published when preparing this text), and hence it is a panel structure.

The firms are studied for their technical efficiency in Section 2 and simultaneously for benefits from further investment in Section 3. The former is expressed with the help of the frontier production function, and the latter is studied using the concept of Tobin´s *Q*.

2 Efficiency

Having a production function $Y = f(K, L)$, we understand the technical efficiency TE_i of the i-th subject as an output-oriented measure defined by the relation

$$TE_i = Y_i / f(K_i, L_i)$$

where Y_i is current output of the subject and $f(K_i, L_i)$ is the feasible technological maximum represented by frontier production function of the group of units to be compared. Evidently, $0 \leq TE_i \leq 1$. Let us suppose the Cobb – Douglas form $Y = AK^{\alpha}L^{\beta}$. Relevant frontier production function can be estimated with the help of the corrected ordinary least squares (COLS) method, which is to be performed in two steps:

[1] Department of Econometrics, University of Economics, Prague, 13067 Praha, nám. Winstona Churchilla 4, Czech Republic, e-mail address: pankova@vse.cz

(i) OLS is used to obtain consistent and unbiased estimates of slope parameters α and β and a consistent but biased estimate of constant $\gamma = \log A$.

(ii) The biased constant $\hat{\gamma}$ is shifted up to encompass all the observed data from above. This is done by setting $\hat{\gamma}' = \hat{\gamma} + max_i\{\hat{u}_i\}$, $\hat{u}_i$ being residuals from the OLS regression.

The production frontier estimated by COLS represents in fact the ´best practice´ technology (for details see e.g. [5])

Now we have $Y_i = \hat{Y}_i \exp(\hat{u}_i)$ and $f(K_i, L_i) = \hat{Y}_i \exp(max_i\{\hat{u}_i\})$. So

$$TE_i = \frac{\hat{Y}_i \exp(\hat{u}_i)}{\hat{Y}_i \exp(max_i\{\hat{u}_i\})} = \exp(\hat{u}_i - max_i\{\hat{u}_i\}). \quad (1)$$

The results are presented as the Figure 1 and in Table 2.

The concept of technical efficiency could be generalized to process panel data, too. Nevertheless, for this part of the analysis, only the 2016 cross-section was used since the results of years 2007 – 2009 approximately are affected not only by the financial crisis but also by different starting procedures in the case of some new firms.

The high differentiation of performance is somewhat surprising. It may partially be explained by a finding that the estimated production function itself has one atypical feature: the role of labour is weaker than the empirical standard. The explanation is a straightforward one: the output and the capital are measured in world prices but the wages in the CR are approximately at the 30 – 40 % level of those in the EU. Incidentally, this is a reason for the presence of many foreign investors in the CR. In any case, managements of particular firms probably consider their economic results as acceptable according to their own internal quantitative criteria.

For a comparison, in the article [4] different branches of the Czech industry were analysed after the first decade of switching from the centrally planned to the market-oriented economy and before a massive inflow of foreign investment. The elasticities of output to capital (about 0.2) and to labour (about 0.8) were the same as in standard economies but the competitiveness of the Czech industry was low.

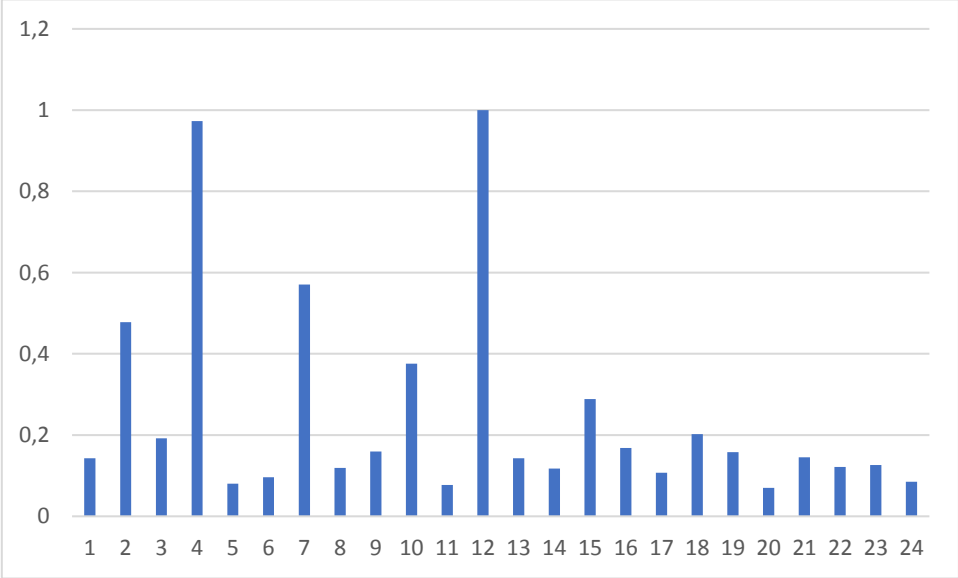


Figure 1 Technical efficiency of firms (No´s of subjects as in Table 1), own computations

3 Tobin´s Q as investment indicator

The investment behaviour of the firm is supposed to be formalized as an optimizing problem

$$V_t \rightarrow MAX$$

subject to

$$I_t - \delta K_t = K_t^*$$

in which V_t is the value of the firm, K_t capital and I_t investment. The solution gives an optimal investment with the goal to maximize the value of the firm. According to Tobin´s theory (see e.g. [3]), investment depends on the ratio Q of relevant shadow price, what in fact is the market value of business capital assets, to their replacement value. The interpretation of such a Q is as follows:

- $Q > 1$ indicates marginal expected profit of a unit of capital to be higher than a unit of additional cost, which is a motivation to invest
- $Q = 1$ is an equilibrium state when there is no incentive for the firm to invest
- $Q < 1$ relates to an unprofitable environment

In practice, neither Q nor the nominator as well as the denominator are observable. A calculation of Q is proposed by Behr and Bellgardt [2] which reflects the fact that only a small part of firms are quoted on stock markets. The Q is computed for firm i at period t using the market value of equity market V_{it}, value of outstanding debt D_{it}, remaining assets aside from the capital stock N_{it} and replacement value of capital stock K_{it} as

$$Q_{it} = \frac{V_{it}+D_{it}-N_{it}}{K_{it}} \quad . \tag{2}$$

All comprised variables are standardly followed except for $V_{it,}$ which can be computed according to the formula

$$\widehat{V_{it}} = \sum_{\tau=1}^{\infty} \widehat{P_{it+\tau}} \delta_{i\tau}^{\tau} \tag{3}$$

with

$$\delta_{i\tau}^{\tau} = \frac{1}{(1+r_{i\tau})^{\tau}}$$

r_{it} being a market interest rate and $\widehat{P_{it}}$ a prediction of pre-tax profits. For applying (2) via (3), a panel data set is proposed. A prediction of $\widehat{P_{it}}$ comes from a VAR model in which sales (S) and cash flows (CF) are comprised.

$$P_{it} = \beta_{10} + \beta_{11}P_{it-1} + \beta_{12}S_{it-1} + \beta_{13}CF_{it-1} + u_{1t}$$

$$S_{it} = \beta_{20} + \beta_{21}P_{it-1} + \beta_{22}S_{it-1} + \beta_{23}CF_{it-1} + u_{2t}$$

$$CF_{it} = \beta_{30} + \beta_{31}P_{it-1} + \beta_{32}S_{it-1} + \beta_{33}CF_{it-1} + u_{3t} \tag{4}$$

In (4), the panel structure of data is important because it provides us with a sufficient number of observations.

This approach can be used universally but for a practical point of view is useful to reflect the finding of Behr and Bellgard [2] that the smaller firms react to Q more strongly. Besides, what is computed by (4) is an average value Q^a while the theory refers to Q as to a marginal value Q^m. In general, it is not $Q^a = Q^m$ (necessary conditions for the equality are formulated and proved by Hayashi [3]). In case of a monopoly, usually $Q^a > Q^m$. The concept of optimal solution is used to formulate an econometric model and to perform a relevant stochastic inference. So, whatever results are found they do not refer neither to global or to local optimal solution of an optimizing problem.

By an application, the question of the number of summands in (3) arises. A common practice (justified in [2]) is to respect the empirically verified rule that it is sufficient to consider four years as a relevant time horizon. The purpose of (4) is not to study its panel structure and characteristics but only to have sufficient data to estimate parameters and to predict $\widehat{P_1} - \widehat{P_4}$. That is why the pool regression was applied after the finding that P (t-value = -5.866), CF (t-value = -8.002) and S (t-value = -10.898) are stationary according to the ADF test (critical t-value = -2.874 at the 5% level).

The average value of Q is 4.473, and the median is 2.988. No result $Q \leq 1$ occurs. To proceed with a description of the Czech industry in the last decade of the past century, the paper [6] could be mentioned dealing with 92 Czech firms (without foreign capital) and showing that the average Q was -0.355, with a median of -0.190.

Further, the Q values are normalized as

$$QN = Q/Q_{MAX}$$

and TE , QN are summarised in Table 1.

No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
TE	0.14	0.47	0.19	0.97	0.08	0.09	0.57	0.11	0.15	0.37	0.07	1
QN	0.05	0.13	0.12	0.81	0.05	0.05	0.22	0.28	0.05	0.20	0.09	1
No	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
TE	0.14	0.11	0.28	0.16	0.10	0.20	0.15	0.07	0.14	0.12	0.12	0.08
QN	0.17	0.04	0.22	0.09	0.12	0.23	0.13	0.06	0.19	0.13	0.04	0.10

Table 1: *TE* and *QN*, own computations

It is Correlation (*TE*, *QN*) = 0.89 what implicates a question about any possible rule what will be a subject of further studies.

5 Conclusions

The sample of an important segment of the Czech industry which operates with the participation of foreign capital is studied for their effectiveness and for profitability of investment.

Contrary to a preliminary expectation, the firms are rather dis-balanced with regard to their production performance, which could be partially explained by their non-standard production function. The role of the labour factor is suppressed due to the cheap labour forces, while capital and output are bought and sold for world prices.

Tobin´s *Q* is defined through an optimization problem what helps to interpret a relation of *Q* to 1 (greater, less or equal) in terms of possible profitability / non-profitability of further investment. The own computations are performed by the help of a panel VAR system.

As for the practice, there is no unprofitable firm in the sample according to Tobin´s *Q* values, It is $Q > 1$ for all units, which means that marginal expected profit of a unit of capital is higher than a unit of additional cost. At least two conclusions result. First, for the firms, it is a motivation to invest. Second, the government should take into consideration whether, in such a positive investment environment, the governmental investment incentives are necessary.

Acknowledgements

Financial support of IGA VŠE F4/78/2018 is gratefully acknowledged.

References

[1] Amadeus32, database, Bureau van Dijk Electronic Publishing.
[2] Behr, A. & Bellgardt, E. (2002). Dynamic Q-investment functions for Germany using panel balance sheet data. *Discussion paper 23* (02), Economic Research Centre of the Deutsche Bundesbank
[3] Hayashi, F. (1982). Tobin´s Marginal Q and Average Q: A neoclassical interpretation, *Econometrica*, Vol. 50, No. 1, pp 213–224
[4] Kodera, J. & Pánková, V. (2001). Capital yields assessment through cross section production function. *Bulletin of Czech Econometric Society*, Vol. 8, No. 14, pp 79–87
[5] Kumbhakar, S. C., Knox Lovell, C.A. (2000*). Stochastic Frontier Analysis*, Cambridge University Press
[6] Pánková, V. (2005). Tobinovo Q – teorie a aplikace. *Politická ekonomie*. Vol. LIII, No. 5, pp 601–608.

Weakly tolerance and possible interval supereigenvectors in fuzzy algebra

Ján Plavka [1]

Abstract. A fuzzy algebra is a nonempty, bounded, linearly ordered set with two binary operations maximum and minimum. A vector x is said to be a λ-eigenvector of a square matrix A if $A \otimes x = \lambda \otimes x$ for some λ. To solve supereigenproblem for some λ means to find a solution x of $A \otimes x \geq \lambda \otimes x$, x is called supereigenvector.
In [1] the properties of supereigenvectors are described, the values λ associated with supereigenvectors are characterized and efficient algorithms for checking all equivalent conditions are introduced.
This contribution investigates the properties of matrices and vectors with interval coefficients. In addition, a complete solution of the weakly tolerance and possible interval supereigenproblem in fuzzy algebra is presented.

Keywords: fuzzy algebra, interval vector, supereigenvector

JEL classification: C44
AMS classification: 15A80, 15A18, 08A72

1 Motivation

The next slightly modified example taken from [6] describes the metropolitan traffic with two trains T_1 and T_2 which run hourly from their home stations S_1 and S_2, respectively, on circular lines which meet in a station S. The trains start their travel in S_1 and S_2, at the times $x_1(0)$ and $x_2(0)$, respectively. In order to allow people to change trains in station S, they must wait 5 minutes after the arrival of the other train, before they can continue. One train needs 20 minutes from S_1 to S, the other train needs 25 minutes from S_2 to S. Let $x_1(r)$ and $x_2(r)$ denote the earliest departure times for train T_1 and train T_2, respectively, from their home stations. Then their schedule develops according to the following system

$$x_1(r+1) = \max(x_1(r) + 20, x_2(r) + 25 + 5) + 20,$$

$$x_2(r+1) = \max(x_1(r) + 20 + 5, x_2(r) + 25) + 25.$$

This system can be written in max-plus algebra as a linear recurrence $x(r+1) = A \otimes x(r) \otimes b$ with

$$A = \begin{pmatrix} 20 & 30 \\ 25 & 25 \end{pmatrix}, \quad b = \begin{pmatrix} 20 \\ 25 \end{pmatrix}.$$

The question in controlling such systems is: How must the system be set in motion to ensure that it moves forward in at least regular steps? This question asks for a constant λ such that $x(r+1) = \lambda \otimes x(r)$ or in vector-matrix notation we have $A \otimes x(r) \geq \lambda \otimes x(r)$, where operation $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$ are extended to matrices and vectors in the same way as in conventional linear algebra.

For simplicity A does not change from stage to stage. The orbit $x,\ A \otimes x, \ldots, A^k \otimes x$, where $A^k = A \otimes \ldots \otimes A$, represents the evolution of such a system. If instead of the equality in $A \otimes x = \lambda \otimes x$ we ask for inequality $\geq$, then we speak about a supereigenproblem.

For a given matrix $A \in \mathbb{B}(n, n)$ in fuzzy algebra ($a \oplus b = \max(a, b)$ and $a \otimes b = \min(a, b)$), the eigenproblem consists of finding a value $\lambda \in \mathbb{B}$ and a vector $x \in \mathbb{B}(n)$ such that the equation $A \otimes x = \lambda \otimes x$ holds true. The eigenproblem in fuzzy algebra has been studied by many authors. Interesting results were found in describing the structure of the eigenspace (the set of all eigenvectors) [19] and several algorithms for computing the largest eigenvector of a given matrix have been suggested, see for example [18].

A vector x is said to be a λ-eigenvector of a square matrix A if $A \otimes x = \lambda \otimes x$ for some $\lambda \in \mathbb{B}$. To solve supereigenproblem for some $\lambda \in \mathbb{B}$ means to find a solution x of $A \otimes x \geq \lambda \otimes x$, x is called λ-supereigenvector. In [17] the properties of λ-supereigenvectors are described, the values λ associated with λ-supereigenvectors are characterized and efficient algorithms for checking all equivalent conditions are introduced.

[1]Technical University in Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

In the present paper, we consider an interval version of this condition and investigate the properties of matrices and vectors with interval coefficients, i.e., we describe λ-supereigenvectors for interval matrices $\boldsymbol{A} = \left\{ A \in \mathbb{B}(n,n);\ \underline{A} \leq A \leq \overline{A} \right\}$ and an interval $\boldsymbol{X} = [\underline{x}, \overline{x}] := \{x \in \mathbb{B}(n);\ \underline{x} \leq x \leq \overline{x}\}$. In addition, a complete solution of the weakly tolerance and the possible interval supereigenproblem in fuzzy algebra is presented.

Let us now give more details on the organization of the paper and on the results obtained there. The next section will be occupied by some definitions and notation of the fuzzy algebra, leading to the discussion of structures of supereigenproblem in Section 3. Section 4 is devoted to the main results of the paper which characterizes interval matrices with the weakly tolerance and the possible interval λ-supereigenvectors.

2 Background of the problem

Let $(\mathbb{B}, \leq)$ be a bounded linearly ordered set with the least element in $\mathbb{B}$ denoted by O and the greatest one by I. The set of naturals (naturals with zero) is denoted by $\mathbb{N}$ ($\mathbb{N}_0$). For given naturals $n, m \in \mathbb{N}$, we use notations N and M for the set of all smaller or equal natural numbers, i.e., $N = \{1, 2, \ldots, n\}$ and $M = \{1, 2, \ldots, m\}$, respectively. The set of $n \times m$ matrices over $\mathbb{B}$ is denoted by $\mathbb{B}(n,m)$, specially the set of $n \times 1$ vectors over $\mathbb{B}$ is denoted by $\mathbb{B}(n)$ and for $\alpha \in \mathbb{B}$ a constant vector is denoted by $\alpha^\star = (\alpha, \ldots, \alpha)^T$.

A fuzzy algebra is a triple $(\mathbb{B}, \oplus, \otimes)$, where $a \oplus b = \max(a,b)$ and $a \otimes b = \min(a,b)$.

The operations $\oplus, \otimes$ are extended to the matrix-vector algebra over $\mathbb{B}$ by the direct analogy to the conventional linear algebra.

For $A \in \mathbb{B}(n,n)$, $C \in \mathbb{B}(n,n)$ we write $A \leq C$ ($A < C$) if $a_{ij} \leq c_{ij}$ ($a_{ij} < c_{ij}$) holds true for all $i, j \in N$. Similarly, for $x = (x_1, \ldots, x_n)^T \in \mathbb{B}(n)$ and $y = (y_1, \ldots, y_n)^T \in \mathbb{B}(n)$ we write $x \leq y$ ($x < y$) if $x_i \leq y_i$ ($x_i < y_i$) for each $i \in N$.

The rth power of a matrix A is denoted by A^r with elements a_{ij}^r.

The $eigenproblem$ in fuzzy algebra is formulated as follows: Given $A \in \mathbb{B}(n,n)$ and the number $\lambda \in \mathbb{B}$, find $x \in \mathbb{B}(n)$ satisfying $A \otimes x = \lambda \otimes x$, n-tuple $x \in \mathbb{B}(n)$ is called *eigenvector* of A and $\lambda \in \mathbb{B}$ is *eigenvalue* of A.

The *eigenspace* $V(A, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e., $V(A,\lambda) = \{x \in \mathbb{B}(n);\ A \otimes x = \lambda \otimes x\}$.

In case $\lambda = I$ let us denote $V(A, I)$ by the abbreviation $V(A)$.

Notice that if $x \in V(A)$, i.e., $A \otimes x = x$, then for each $\lambda \in \mathbb{B}$ the equalities

$$A \otimes (\lambda \otimes x) = \lambda \otimes (A \otimes x) = \lambda \otimes x = \lambda \otimes (\lambda \otimes x)$$

hold true. Thus, every $\lambda \in \mathbb{B}$ is an eigenvalue of A.

For every matrix $A \in \mathbb{B}(n,n)$ denote

$$c_i(A) = \bigoplus_{j \in N} a_{ij},\ c(A) = \bigotimes_{i \in N} c_i(A),\ c_A = (c_1(A), c_2(A), \ldots, c_n(A))^T \in \mathbb{B}(n),$$

$$\underline{\lambda}(A)(= \underline{\lambda}) = \bigotimes_{i \in N} a_{ii} \ \text{ and } \overline{\lambda}(A)(= \overline{\lambda}) = \bigoplus_{i,j \in N} a_{ij}\ .$$

Notice that both operations in fuzzy algebra are idempotent, hence the range of mappings $\max$ and $\min$ is $\mathbb{B}$ and no new numbers are created in the process of generating matrix-vector products.

3 Structures of supereigenvectors

For $A \in \mathbb{B}(n,n)$ and $\lambda \in \mathbb{B}$ we denote $V^*(A,\lambda) = \{x \in \mathbb{B}(n); A \otimes x \geq \lambda \otimes x\}$ or just $V^*(A)$ for $\lambda = I$.
Notice that the sets $V^*(A,\lambda)$ is not empty for each A and for each λ because there is a greatest eigenvector belonging to $V^*(A,\lambda)$.

If $A \in \mathbb{B}(n,n)$ and $\lambda \in \mathbb{B}$ then a vector $x \in \mathbb{B}(n)$ satisfying $A \otimes x \geq \lambda \otimes x$ is called a λ-supereigenvector of A with associated supereigenvalue λ and $V^*(A,\lambda)$ is called λ-supereigenspace.

Lemma 1. [17] Let $A \in \mathbb{B}(n,n)$ be given. The following implications hold true:
(i) If $\underline{\lambda} \leq \lambda \leq c(A)$ and $x \geq c^\star(A)$, then $x \in V^*(A,\lambda)$.

(ii) If $\alpha \le c(A)$, then $(\forall\lambda)[\alpha^\star \in V^*(A,\lambda)]$.
(iii)If $c(A) \le \lambda \wedge x \le c^\star(A)$ or $\overline{\lambda} \le \lambda \wedge c^\star(A) \le x \le c_A$, then $V^*(A,\lambda) = V^*(A,I)$.
(iv)If $c(A) < \lambda \ \wedge \ x \ge \overline{\lambda}^\star \ \wedge \ x > c^\star(A)$, then $V^*(A,\lambda) = \emptyset$.
(v) If $\lambda \in \mathbb{B}$, $c_A \le x \le \overline{\lambda}^\star$ and $x \in V^*(A,\lambda)$, then $\lambda \otimes x = \lambda \otimes c_A$.

For $A \in \mathbb{B}(n,n)$ denote $\bigotimes_{(A\otimes c_A)_i < (c_A)_i} (A \otimes c_A)_i$ by $\tilde{\lambda}(A)$ or just $\tilde{\lambda}$.

Theorem 2. [17] *If $c_A \le x \le \overline{\lambda}^\star$, then the following hold true:*
(i) If $\lambda \le \tilde{\lambda}$, then $x \in V^*(A,\lambda) \Leftrightarrow \lambda \otimes x = \lambda \otimes c_A$.
(ii) If $\tilde{\lambda} < \lambda$, then $V^*(A,\lambda) = \emptyset$.

Lemma 3. [17] Let $A \in \mathbb{B}(n,n)$ and $\lambda \in \mathbb{B}$ be given. Then there exists an algorithm which correctly decides whether $x \in V^*(A,\lambda)$ in $O(n^2)$ arithmetic operations.

4 Interval versions of λ-supereigenvectors

Similarly to [7], [11], [12], [13], [16], we define interval matrix with bounds $\underline{A}, \overline{A} \in \mathbb{B}(n,n)$ and interval vector with bounds $\underline{x}, \overline{x} \in \mathbb{B}(n)$ as follows

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathbb{B}(n,n);\ \underline{A} \le A \le \overline{A} \right\}, \quad \boldsymbol{X} = [\underline{x}, \overline{x}] = \left\{ x \in \mathbb{B}(n);\ \underline{x} \le x \le \overline{x} \right\}.$$

We assume in this section that an interval matrix $\boldsymbol{A} = [\underline{A}, \overline{A}]$, $\lambda \in \mathbb{B}$ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are fixed. The interval supereigenproblem for $\boldsymbol{A}$, $\lambda \in \mathbb{B}$ and $\boldsymbol{X}$ consists in recognizing whether $A \otimes x \ge \lambda \otimes x$ holds true for $A \in \boldsymbol{A}$, $x \in \boldsymbol{X}$. In dependence on the applied quantifiers, we get two types of interval λ-supereigenvectors.

Definition 1. If interval matrix $\boldsymbol{A}$ and $\lambda \in \mathbb{B}$ are given, then interval vector $\boldsymbol{X}$ is called
- *possible λ-supereigenvector* of $\boldsymbol{A}$ if $(\exists A \in \boldsymbol{A})(\forall x \in \boldsymbol{X})[\, A \otimes x \ge \lambda \otimes x \,]$,
- *weakly tolerance λ-supereigenvector* of $\boldsymbol{A}$ if $(\exists A \in \boldsymbol{A})(\exists x \in \boldsymbol{X})\, [A \otimes x \ge \lambda \otimes x \,]$.

The remainder of this paper assumes we are given an interval matrix $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$. For each pair of indices $i, j \in N$, we define $\tilde{A}^{(ij)} \in \mathbb{B}(n,n)$ and $\tilde{x}^{(i)} \in \mathbb{B}(n)$ by putting for every $k, l \in N$,

$$\tilde{a}_{kl}^{(ij)} = \begin{cases} \overline{a}_{ij}, & \text{for } k = i,\ l = j \\ \underline{a}_{kl}, & \text{otherwise} \end{cases}, \qquad \tilde{x}_k^{(i)} = \begin{cases} \overline{x}_i, & \text{for } k = i \\ \underline{x}_k, & \text{otherwise} \end{cases}.$$

The next lemma says that every matrix in $\boldsymbol{A}$ can be written as a max-min linear combination of *generating* matrices ("generators," for short) $\tilde{A}^{(ij)}$ with $i,\ j \in N$. Similarly, every vector in $\boldsymbol{X}$ is equal to a max-min linear combination of generators $\tilde{x}^{(i)}$ with $i \in N$.

Lemma 4. Let $x \in \boldsymbol{X}$ and $A \in \boldsymbol{A}$. Then $x = \bigoplus_{i=1}^{n} \beta_i \otimes \tilde{x}^{(i)}$ for some numbers $\beta_i \in \mathbb{B}$ and $A = \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ for some numbers $\alpha_{ij} \in \mathbb{B}$.

Proof. Suppose that $x \in \boldsymbol{X}$ and $A \in \boldsymbol{A}$. It is easily seen that the assertions follow for $\beta_i = x_i$ and $\alpha_{ij} = a_{ij}$. □

4.1 Possible λ-supereigenvectors

In the next theorems we present the equivalent conditions for checking if a given vector is possible λ-supereigenvector of $\boldsymbol{A}$.

Theorem 5. *Let $\boldsymbol{A}$, $\boldsymbol{X}$ and $\lambda \in \mathbb{B}$ be given. The interval vector $\boldsymbol{X}$ is a possible λ-supereigenvector of $\boldsymbol{A}$ if and only if $(\exists A \in \boldsymbol{A})(\forall k \in N)[A \otimes \tilde{x}^{(k)} \ge \lambda \otimes \tilde{x}^{(k)}]$.*

Proof. Assume that there is $\lambda \in \mathbb{B}$ and $A \in \boldsymbol{A}$ such that $(\forall k \in N)[A \otimes \tilde{x}^{(k)} \ge \lambda \otimes \tilde{x}^{(k)}]$. Then by Lemma 4(i) for arbitrary $x \in \boldsymbol{X}$ we get

$$A \otimes x = A \otimes \bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)} = \bigoplus_{k=1}^{n} \beta_k \otimes (A \otimes \tilde{x}^{(k)}) \ge \bigoplus_{k=1}^{n} \beta_k \otimes (\lambda \otimes \tilde{x}^{(k)}) = \lambda \otimes x.$$

The converse implication trivially follows. □

Let $\boldsymbol{A}$ and $\boldsymbol{X}$ be given. Denote the block matrix $\tilde{A} \in B(n^2, n^2)$ and vectors $\tilde{x}, \alpha \in B(n^2)$ as follows

$$\tilde{A} = \begin{pmatrix} A^{(11)} \otimes x^{(1)} & \dots & A^{(1n)} \otimes x^{(1)} & A^{(21)} \otimes x^{(1)} & \dots & A^{(nn)} \otimes x^{(1)} \\ A^{(11)} \otimes x^{(2)} & \dots & A^{(1n)} \otimes x^{(2)} & A^{(21)} \otimes x^{(2)} & \dots & A^{(nn)} \otimes x^{(2)} \\ \vdots & & & & & \\ A^{(11)} \otimes x^{(n)} & \dots & A^{(1n)} \otimes x^{(n)} & A^{(21)} \otimes x^{(n)} & \dots & A^{(nn)} \otimes x^{(n)} \end{pmatrix},$$

$$\tilde{x} = (x_1^{(1)}, \dots, x_n^{(1)}, x_1^{(2)}, \dots, x_n^{(2)}, \dots, x_1^{(n)} \dots, x_n^{(n)})^T,$$

and

$$\alpha = (\alpha_{11}, \dots, \alpha_{1n}, \alpha_{21}, \dots, \alpha_{2n}, \dots, \alpha_{n1}, \dots, \alpha_{nn})^T.$$

Theorem 6. *Let $\boldsymbol{A}$ be an interval matrix and let $\boldsymbol{X}$ be an interval vector. Then the interval vector $\boldsymbol{X}$ is a possible λ-supereigenvector of $\boldsymbol{A}$ if and only if the system $\tilde{A} \otimes \alpha \geq \tilde{x}$ has a solution α such that $\underline{A} \leq \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \leq \overline{A}$.*

Proof. Suppose that the system $\tilde{A} \otimes \alpha \geq \tilde{x}$ is solvable, i.e., there is a vector α such that

$$\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \otimes x^{(k)} \geq x^{(k)} \text{ for all } k \in N$$

and $\underline{A} \leq \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \leq \overline{A}$. Put $A = \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)}$. Then for each $x = \bigoplus_{i=1}^{n} \beta_i \otimes x^{(i)}$ from $\boldsymbol{X}$ we get

$$A \otimes x = (\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)}) \otimes (\bigoplus_{k=1}^{n} \beta_k \otimes x^{(k)}) =$$

$$\bigoplus_{k=1}^{n} \beta_k \otimes \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \otimes x^{(k)} \geq \bigoplus_{k=1}^{n} \beta_k \otimes x^{(k)} = x.$$

By Theorem 5 the assertion holds true. □

Theorem 7. [2] *Let $C, D \in B(m, n)$ and $e, f \in B(m)$. Then the system of inequalities $C \otimes x \leq e$, $D \otimes x \geq f$ has a solution if and only if $D \otimes \hat{x}(C, e) \geq f$, where $\hat{x}_j(C, e) = \min_{i \in M} \{e_i : c_{ij} > e_i\}$ for $j \in N$ ($\min \emptyset = I$).*

Observe that the solvability of the system of inequalities $C \otimes x \leq e$, $D \otimes x \geq f$ can be check in $O(mn)$ elementary operations.

The last theorem enables us to check the sufficient condition of Theorem 6 in practice, whereby $\tilde{A} \otimes \alpha \geq \tilde{x}$ and $\underline{A} \leq \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \leq \overline{A}$ are joint into two systems of inequalities according to Theorem 7 as follows:

$$\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \leq \overline{A} \quad \text{and} \quad \begin{matrix} \tilde{A} \otimes \alpha \geq \tilde{x}, \\ \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes A^{(ij)} \geq \underline{A}. \end{matrix}$$

Denote i-th row and j-th column of A by $A_{i\,\cdot}$ and $A_{\cdot\,j}$, respectively.

Let $\boldsymbol{A}$ and $\boldsymbol{X}$ be given. Denote block matrices $C \in B(n^2, n^2)$, $D \in B(2n^2, n^2)$ and vectors $e \in B(n^2)$, $f \in B(2n^2)$ as follows

$$C = \begin{pmatrix} A_{\cdot 1}^{(11)} & \dots & A_{\cdot 1}^{(1n)} & A_{\cdot 1}^{(21)} & \dots & A_{\cdot 1}^{(nn)} \\ \vdots & & & & & \\ A_{\cdot n}^{(11)} & \dots & A_{\cdot n}^{(1n)} & A_{\cdot n}^{(21)} & \dots & A_{\cdot n}^{(nn)} \end{pmatrix}, \; e = \begin{pmatrix} \overline{A}_{\cdot\,1} \\ \overline{A}_{\cdot\,2} \\ \vdots \\ \overline{A}_{\cdot\,n} \end{pmatrix},$$

$$D = \begin{pmatrix} A^{(11)} \otimes x^{(1)} & \dots & A^{(1n)} \otimes x^{(1)} & A^{(21)} \otimes x^{(1)} & \dots & A^{(nn)} \otimes x^{(1)} \\ A^{(11)} \otimes x^{(2)} & \dots & A^{(1n)} \otimes x^{(2)} & A^{(21)} \otimes x^{(2)} & \dots & A^{(nn)} \otimes x^{(2)} \\ \vdots & & & & & \\ A^{(11)} \otimes x^{(n)} & \dots & A^{(1n)} \otimes x^{(n)} & A^{(21)} \otimes x^{(n)} & \dots & A^{(nn)} \otimes x^{(n)} \\ A_{\cdot 1}^{(11)} & \dots & A_{\cdot 1}^{(1n)} & A_{\cdot 1}^{(21)} & \dots & A_{\cdot 1}^{(nn)} \\ \vdots & & & & & \\ A_{\cdot n}^{(11)} & \dots & A_{\cdot n}^{(1n)} & A_{\cdot n}^{(21)} & \dots & A_{\cdot n}^{(nn)} \end{pmatrix}, \quad f = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \\ \underline{A}_{\cdot 1} \\ \vdots \\ \underline{A}_{\cdot n} \end{pmatrix}$$

and

$$\alpha = (\alpha_{11}, \dots, \alpha_{1n}, \alpha_{21}, \dots, \alpha_{2n}, \dots, \alpha_{n1}, \dots, \alpha_{nn})^T.$$

Theorem 8. *Let $\boldsymbol{A}$ be an interval matrix and let $\boldsymbol{X}$ be an interval vector. Then the interval vector $\boldsymbol{X}$ is a possible λ-supereigenvector of $\boldsymbol{A}$ if and only if the system of inequalities $C \otimes \alpha \leq e$, $D \otimes \alpha \geq f$ is solvable.*

Proof. The assertion follows from Theorem 6 and Theorem 7. □

Theorem 9. *Suppose given a matrix $\boldsymbol{A}$, λ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$. The recognition problem of whether $\boldsymbol{X}$ is possible λ-supereigenproblem of $\boldsymbol{A}$ is solvable in $O(n^3)$ time.*

Proof. According to Theorem 8, the recognition problem whether $\boldsymbol{X}$ is possible λ-supereigenproblem is equivalent to recognizing whether the system of inequalities $C \otimes \alpha \leq e$, $D \otimes \alpha \geq f$ is solvable. The computation of the system of inequalities $C \otimes \alpha \leq e$, $D \otimes \alpha \geq f$ needs $O(2n^2 \cdot n) = O(n^3)$ time (see [2]). □

4.2 Weakly tolerance λ-supereigenvectors

Denote the block matrices $\tilde{B} \in \mathbb{B}(n, n^3), \tilde{B}^{(k)} \in \mathbb{B}(n, n^2), \tilde{C} \in \mathbb{B}(n, n)$ and vectors $y \in \mathbb{B}(n^3), \alpha \in \mathbb{B}(n^2), \beta \in \mathbb{B}(n)$ as follows

$$\tilde{B}^{(k)} = \begin{pmatrix} \tilde{A}^{(11)} \otimes \tilde{x}^{(k)} & \dots & \tilde{A}^{(1n)} \otimes \tilde{x}^{(k)} & \tilde{A}^{(21)} \otimes \tilde{x}^{(k)} & \dots & \tilde{A}^{(nn)} \otimes \tilde{x}^{(k)} \end{pmatrix},$$

$$\tilde{B} = \begin{pmatrix} \tilde{B}^{(1)} & \tilde{B}^{(2)} & \dots & \tilde{B}^{(n)} \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} \tilde{x}^{(1)} \tilde{x}^{(2)} \dots \tilde{x}^{(n)} \end{pmatrix},$$

and

$$y = (y_{111}, \dots, y_{1n1}, y_{211}, \dots, y_{2n1}, \dots, y_{nn1}, y_{112}, \dots, y_{nnn})^T,$$
$$\alpha = (\alpha_{11}, \dots, \alpha_{1n}, \alpha_{21}, \dots, \alpha_{2n}, \dots, \alpha_{n1}, \dots, \alpha_{nn})^T, \quad \beta = (\beta_1, \beta_2, \dots, \beta_n)^T.$$

We say that the quadratic two-sided linear system of inequalities $\tilde{B} \otimes y \geq \lambda \otimes \tilde{C} \otimes \beta$ is weakly solvable if there exist vectors y, α, β such that $\tilde{B} \otimes y \geq \lambda \otimes \tilde{C} \otimes \beta$ and $\alpha_{ij} \otimes \beta_k = y_{ijk}$ for all $i, j, k \in N$.

In the next theorem we present the equivalent conditions for a given vector to be weakly tolerance λ-supereigenvector of $\boldsymbol{A}$.

Theorem 10. *Let $\boldsymbol{A}$, $\boldsymbol{X}$ and λ be given. Then $\boldsymbol{X}$ is weakly tolerance λ-supereigenvector of $\boldsymbol{A}$ if and only if the quadratic two-sided linear system $\tilde{B} \otimes y \geq \lambda \otimes \tilde{C} \otimes \beta$ is weakly solvable.*

Proof. Suppose that $\boldsymbol{X}$ is weakly tolerance λ-supereigenvector of $\boldsymbol{A}$. Then, by the definition, $\boldsymbol{X}$ is weakly tolerance λ-supereigenvector of $\boldsymbol{A}$ if $(\exists A \in \boldsymbol{A})(\exists x \in \boldsymbol{X})\, [A \otimes x \geq \lambda \otimes x\,]$. In view of Lemma 4, we have

$$(\exists A \in \boldsymbol{A})(\exists x \in \boldsymbol{X}\,)[A \otimes x \geq \lambda \otimes x] \Leftrightarrow$$

$$(\exists \alpha_{ij})(\exists \beta_k)(\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes \tilde{A}^{(ij)}) \otimes (\bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)}) \geq \lambda \otimes \bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)} \Leftrightarrow$$

$$(\exists \alpha_{ij})(\exists \beta_k) \bigoplus_{k=1}^{n} \bigoplus_{i,j=1}^{n} (\alpha_{ij} \otimes \tilde{A}^{(ij)}) \otimes (\beta_k \otimes \tilde{x}^{(k)}) \geq \lambda \otimes \bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)} \Leftrightarrow$$

$$(\exists \alpha_{ij})(\exists \beta_k) \bigoplus_{k=1}^{n} \bigoplus_{i,j=1}^{n} (\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)}) \otimes (\alpha_{ij} \otimes \beta_k) \geq \lambda \otimes \bigoplus_{k=1}^{n} \tilde{x}^{(k)} \otimes \beta_k \Leftrightarrow$$

$$(\exists\alpha)(\exists\beta)(\exists y)(\tilde{B}\otimes y\geq\lambda\otimes\tilde{C}\otimes\beta) \text{ and } \alpha_{ij}\otimes\beta_k=y_{ijk}\ \forall i,j,k\in N,$$

i.e., the quadratic two-sided linear system $\tilde{B}\otimes y\geq\lambda\otimes\tilde{C}\otimes\beta$ is weakly solvable.

To prove the reverse implication, assume that the system $\tilde{B}\otimes y\geq\lambda\otimes\tilde{C}\otimes\beta$ is weakly solvable. As we have shown above the numbers α_{ij} and β_k for all $i,j,k\in N$ can be computed from the weak solution of the system $\tilde{B}\otimes y\geq\lambda\otimes\tilde{C}\otimes\beta$ and hence there exists a matrix $A\in\boldsymbol{A}$ and a vector $x\in\boldsymbol{X}$ such that $A\otimes x\geq\lambda\otimes x$. □

Notice that Theorem 10 implies the computational complexity of a procedure based on checking all solutions can be exponentially large. We are able neither to suggest polynomial algorithm nor to prove NP-hardness of weak tolerance λ-supereigenvector of $\boldsymbol{X}$.

References

[1] Butkovič, P.: Max-linear Systems: Theory and Algorithms, *Springer Monographs in Mathematics*, Springer-Verlag 2010.

[2] K. Cechlárová: Solutions of Interval Linear Systems in $(max,+)$-algebra. In: Proceedings of the 6th International Symposium on Operational Research Preddvor, Slovenia 2001, 321–326.

[3] Cuninghame-Green, R. A.: Minimax Algebra, *Lecture Notes in Economics and Mathematical Systems* vol. 166, Springer, Berlin, 1979.

[4] Gavalec, M.: Periodicity in Extremal Algebra, Gaudeamus, Hradec Králové 2004.

[5] Gavalec, M., Gad, M. and Zimmermann, K.: Optimization Problems under (max, min)- Linear Equations and / or Inequality Constraints, Journal of Mathematical Sciences 193, 5 (2013), 645–658.

[6] B. Heidergott, G. J. Olsder and J. van der Woude, Max plus at work. Modeling and analysis of synchronized systems. Princeto University Press, 2004.

[7] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *Linear Algebra and Its Applications* 438(8) (2013) 3350–3364.

[8] Molnárová, M.: Generalized matrix period in max-plus algebra, *Linear Algebra and Its Applications* 404 (2005) 345–366.

[9] Molnárová, M.: Computational complexity of Nachtigall's representation, *Optimization* 52(1) (2003) 93–104.

[10] Molnárová, M. and Pribiš, J.: Matrix period in max-algebra, *Discrete Applied Mathematics* 103(1-3) (2000) 167–175.

[11] Myšková, H.: Robustness of Interval Toeplitz Matrices in Fuzzy Algebra, Acta Electrotechnica et Informatica, Vol. 12, No. 4, 2012, 56–60.

[12] Myšková, H.: Interval eigenvectors of circulant matrices in fuzzy algebra *Acta Electrotechnica et Informatica*. Vol. 12, No. 3, 2012, 57–61.

[13] Myšková, H.: Weak stability of interval orbits of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica*, Vol. 12, No. 3, 2012, 51–56.

[14] Plavka J., and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *Discrete Applied Mathematics* 159(5) (2011), 381–388.

[15] Plavka, J.: On the weak robustness of fuzzy matrices, *Kybernetika* 49(1) (2013), 128–140.

[16] Plavka, J., and Sergeev, S.: Characterizing matrices with X-simple image eigenspace in max-min semiring, *Kybernetika* 52(4) (2016), 497–513.

[17] Plavka, J.: Subeigenvectors and supereigenvectors of fuzzy matrices, Fuzzy Sets and Systems 339 (2018) 62–72.

[18] Tan, Yi-Jia: On the eigenproblem of matrices over distributive lattices, *Linear Algebra and its Applications* 374 (2003), 87–106.

[19] Tomášková, H: Eigenproblem for circulant matrices in max-plus algebra. In: Proceedings of the 29th Conference on Mathematical Methods, Computational Techniques, Intelligent Systems (MAMECTIS-29), 2010.

[20] Zimmernann, K.: Extremální algebra (in Czech), *Ekon. ústav ČSAV* Praha, 1976.

Estimating the Laffer Curve for Slovakia: A DSGE Approach

Daniel Němec[1], Vlastimil Reichel[2], Jakub Chalmovianský[3], Jana Balážová[4]

Abstract. The goal of this paper is to quantify the impact of alternative tax adjustments on government tax revenues and the size of the shadow economy in Slovakia. Our approach uses a small dynamic stochastic general equilibrium (DSGE) model consisting of both formal and informal sectors of the economy. Using the quarterly data for the Slovak economy from 2000 to 2017 we estimate a linearised form of the DSGE model. Based on our parameter estimates, we perform a set of simulations using the non-linear form of the underlying DSGE model to evaluate the effects of changes in the corporate tax rate, personal income tax rate, social security tax rate, and in the probability of tax control. These effects are evaluated with regards to the changes in steady-state values of the output in the official economy, total tax revenues, and the size of the shadow economy. The corresponding Laffer curves show strong responses in the size of the underground economy when the corporate tax rates or personal income tax rates are growing. On the other hand, the increase of the social security tax rates causes only small growth in the share of the shadow economy without negative effects on government tax revenues.

Keywords: underground economy, DSGE model, tax system, Laffer curve, Slovakia.

JEL Classification: C11, C53, H26
AMS Classification: 91B64,

1 Introduction

According to Feld et al. [1], the underground economy is part of a non-observed economy (NOE), which contains all unregistered economic activities and the production of goods and services on the market, whether they are legal or illegal. An important aspect is that these activities are not included in official GDP estimates. In 2002, the Organization for Economic Operations and Development (OECD) published a handbook on how countries (which are part of the OECD) should count NOE in their national accounts. According to the OECD methodology, NOE is divided into four groups: underground economy, illegal activities, informal sector and unpaid household economic activities for its own benefit (OECD, [3]).

Monitoring the evolution or measuring the size of the NOE is intrinsically complex. Therefore, it is only possible to estimate its components. There are three main methods to measure the size of the non-observed economy. The first method uses a direct approach. The estimation of the unobservable economy in this approach is based on questionnaire surveys and their subsequent evaluation (Feld et al., [1]; Kirchgassner, [2]). The second method uses indirect approaches. Feld et al. [1] and Schneider and Buhn [7] argue that these approaches include (i) estimating the unobserved economy as a difference between national income and expenditure, (ii) the difference between official and actual labour, (iii) transaction approach and (iv) the demand for currency approach. The third method of assessing the size of an unobservable economy is through a model where the unobservable economy is viewed as a hidden variable. The best-known and most widely used model is the Multiple Indicator Multiple Cause (MIMIC) model. The second type of model used in this method is the dynamic stochastic general equilibrium (DSGE) model.

The main goal of this paper is to quantify the impact of alternative tax adjustments on government tax revenues and the size of the shadow economy in Slovakia. For this purpose, we use a estimated DSGE model that is based on the model proposed by Orsi et al. [4]. We use this particular model because it consists of both formal and informal sector and it suits the dynamics of the Slovak economy quite well. The results of our set of simulations of the non-linear form of the underlying DSGE model correspond to Laffer curves. Another example of a DSGE

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, daniel.nemec@econ.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, reichel.v@mail.muni.cz
[3] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, chalmoviansky@mail.muni.cz
[4] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, j.balazova@mail.muni.cz

model used to obtain a Laffer curve is Solis-Garcia et al. [8]. Based on the obtained Laffer curves, we can learn about the incentive effects of tax changes and the degree of self-financing of tax cuts (Trabandt and Uhlig [9]).

2 Model

In this section, we present the model of the NOE based on Orsi et al. [4]. The model consists of three sectors: households, firms and government. Firms employ an official workforce $h^m_{i,t}$, and hire official capital $k^m_{i,t}$. However, some companies are trying to hide part of their production to avoid taxation. In this case, these companies employ an unofficial workforce h^u_t and use unofficial capital k^u_t. The homogeneous output of all companies is defined using the Cobb-Douglas production function (1):

$$y^j_{i,t} = A^j_t(\Gamma_t h^j_{i,t})^{\alpha_j}(k^j_{i,t})^{1-\alpha_j}, \quad j = \{m, u\}; \qquad y_{i,t} = y^m_{i,t} + y^u_{i,t}, \tag{1}$$

where $\alpha_j \in (0, 1)$. A^j_t is a technological shock and Γ_t is the labour-augmenting technological progress that follows a deterministic trend $\Gamma_t = \gamma\Gamma_{t-1}$, where $\gamma > 1$.

Capital and labour are hired in a perfectly competitive market. Companies pay rent, r^m_t or r^u_t, per unit of capital and wage, w^m_t or w^u_t, per unit of work, whether or not they produce in the NOE. In addition, the official wage, w^m_t, is increased by the social and health insurance rate paid by the company, τ^s_t. Thus, the total cost can be defined as follows:

$$TC(h^m_{i,t}, h^u_{i,t}, k^m_{i,t}, k^u_{i,t}) = (1 + \tau^s_t)w^m_t h^m_{i,t} + w^u_t h^u_{i,t} + r^u_t k^u_{i,t} + r^m_t k^m_{i,t}. \tag{2}$$

The government is trying to prevent the shift of production to the NOE and therefore introduces controls. Every company at time t is facing the probability of being subject of a check, $p_t \in (0, 1)$, which may result in an obligation to pay the unpaid tax based on the tax rate, $\tau^c_t < 1$, and the penalty surcharge, $s > 1$. The result is that the net output gain at time t, $y_{i,t}$, is a random variable defined by the expression:

$$NR(y_{i,t}) = \left\{ \begin{array}{l} y^m_{i,t} - \tau^c_t(y^m_{i,t} - w^m_t h^m_{i,t}) + y^u_{i,t} - s\tau^c_t(y^u_{i,t} - w^u_t h^u_{i,t}) \\ y^m_{i,t} - \tau^c_t(y^m_{i,t} - w^m_t h^m_{i,t}) + y^u_{i,t} \end{array} \right\}. \tag{3}$$

A representative household maximises the utility function (4), subject to capital (5) and budget constraint (6).

$$U^h_t = \sum_{t=0}^{\infty} \beta^t E_0 \left\{ \frac{(\frac{c_t}{\Gamma_t})^{1-\sigma} - 1}{1-\sigma} - \xi^h_t B_0 \frac{(h^m_t + h^u_t)^{1+\xi}}{1+\xi} - B_1 \frac{(h^u_t)^{1+\psi}}{1+\psi} \right\}, \tag{4}$$

$$k_{t+1} = \xi^x_t x_t + (1 - \delta_k)k_t, \qquad \text{where } k_t = k^u_t + k^m_t, \tag{5}$$

$$c_t + x_t = (1 - \tau^h_t)(w^m_t h^m_t + r^m_t k^m_t) + w^u_t h^u_t + r^u_t k^u_t. \tag{6}$$

Where $\sigma > 0$ is the inverse of the intertemporal elasticity of substitution; $\beta \in (0, 1)$ is the discount factor; $B_0 \geq 0$, $B_1 \geq 0$ are preference parameters controlling for the disutility of work; $\psi > 0$, $\xi > 0$ denote the inverse elasticities of aggregate and underground labour supplies, respectively; ξ^h_t denotes a transitory labour supply shock that affects the marginal rate of substitution between consumption and leisure; k_t is the capital held by households; x_t indicates the investment at time t; and $\delta_k \in [0, 1]$ denotes the capital depreciation rate.

The government raises taxes to finance government consumption, g_t. For simplicity, we abstract from public debt and assume that the amount of public expenditure is based on a balanced budget. Therefore, the government budget constraint is as follows:

$$g_t = \tau^h_t(w^m_t h^m_t + r^m_t k^m_t) + \tau^c_t \int_0^1 [y^m_{i,t} - w^m_t h^m_{i,t} + p_t s(y^u_{i,t} - w^u_t h^u_{i,t})]\mathrm{d}i + \tau^s_t w^m_{i,t} \int_0^1 h^m_{i,t}\mathrm{d}i, \tag{7}$$

where the first expression on the RHS of the equation indicates the amount of fiscal revenues from personal income taxation, g^h_t, the second term is total fiscal revenues from corporate taxation, g^c_t, and the last term stands for total fiscal revenues from social security contributions, g^s_t. However, at the time t there is also a total tax evasion, TE_t, having the following form:

$$TE_t = \tau^s_t w^m_{i,t} \int_0^1 h^m_{i,t}\mathrm{d}i + \tau^h_t(w^m_t h^m_t + r^m_t k^m_t) + (1 - p_t)\tau^c_t \int_0^1 (y^u_{i,t} - w^u_t h^u_{i,t})\mathrm{d}i. \tag{8}$$

The model consists of eight variables (two technological shocks, investment efficiency shock, labour supply shock, three tax rate shocks, and shock in probability to be investigated) following (in the log-linear form) an AR(1) process (see Table 2 for detailed specification).

3 Data and methodology

We use seven macroeconomic variables for the estimation of the DSGE model, namely: households consumption, investment, government revenue from corporate tax, government revenue from the social and health insurance contributions, government revenue from the taxation of individuals, total gross wages, and the likelihood of corporate tax audits to households. The data are obtained from the Statistical Office of the Slovak Republic (SOSR), the Ministry of Finance of the Slovak Republic (MFSR), or the Financial Administration of Slovak Republic (FASR), covering the period from the 1st quarter of 2000 to the fourth quarter of 2017. We transform variables described above using first logarithmic differences (Δ). In the case of the probability of tax auditing, we recalculated the series using the formula $\tilde{p_t} = \frac{p_t - \bar{p}}{\bar{p}}$, where $\tilde{p_t}$ denotes the transformed time series used in the model, and $\bar{p}$ is the average probability of tax auditing in the examined period (i.e. the steady state value), as suggested in Orsi et al. [4] (see Table 1).

Shortcut	Variable description	Data Source	Transformation
C	Households consumption	SOSR	log, Δ
X	Investment	SOSR	log, Δ
G^c	Government corporate tax revenues	MFSR	log, Δ
G^s	Government revenues from social and health insurance contributions	MFSR	log, Δ
G^h	Government personal income tax revenues	MFSR	log, Δ
W	Total gross wages	SOSR	log, Δ
p	The probability of tax auditing	FASR	$\tilde{p_t} = \frac{p_t - \bar{p}}{\bar{p}}$

Table 1 Data description

		posterior			prior		
parameters		mean	90% HPDI		mean	std. dev.	distribution
α_m	share of the total number of hours worked in OE	0.6016	[0.5664	0.6388]	0.66	0.02	beta
α_u	share of the total number of hours worked in UE	0.6555	[0.6229	0.6895]	0.66	0.02	beta
σ	inverse of the intertemporal elasticity of substitution	1.7661	[0.8043	2.7129]	1.80	4.20	gamma
ξ	inverse labour supply elasticities of aggregate labour supply	1.6353	[1.4162	1.8561]	1.49	0.13	gamma
ψ	inverse labour supply elasticities of underground labour supply	1.0004	[0.9674	1.0332]	1	0.02	gamma
δ_k	capital depreciation rate	0.0236	[0.0042	0.0416]	0.029	0.016	beta
B_1	disutility of working activities	6,1216	[4.4708	7.8291]	6	1	gamma
persistence of the shocks							
ρ_{a^m}	persistence of technology shock in OE	0.1400	[0.0445	0.2277]	0.5	0.15	beta
ρ_{a^u}	persistence of technology shock in NOE	0.1378	[0.0495	0.2245]	0.5	0.15	beta
ρ_{τ^c}	persistence of tax rate shock	0.1345	[0.0444	0.2199]	0.5	0.15	beta
ρ_{τ^s}	persistence of social security tax rate shock	0.1425	[0.0461	0.2327]	0.5	0.15	beta
ρ_{τ^h}	persistence of income tax rate shock	0.1455	[0.0525	0.2358]	0.5	0.15	beta
ρ_{ξ^x}	persistence of purely transitory exogenous shock	0.1421	[0.0496	0.2305]	0.5	0.15	beta
ρ_{ξ^h}	persistence of labour supply shock	0.4847	[0.2473	0.7332]	0.5	0.15	beta
ρ_p	persistence of probability of tax audit shock	0.0314	[0.0078	0.0543]	0.5	0.15	beta
standard deviation of shocks							
σ_{a^m}	std. dev. of technology shock in OE	0,1042	[0.0866	0.1204]	0.05	∞	IG
σ_{a^u}	std. dev. of technology shock in NOE	1.0134	[0.8554	1.1688]	0.05	∞	IG
σ_{τ^c}	std. dev. of tax rate shock	0.0517	[0.0441	0.0594]	0.05	∞	IG
σ_{τ^s}	std. dev. of social security tax rate shock	0.0145	[0.0125	0.0167]	0.05	∞	IG
σ_{τ^h}	std. dev. of income tax rate shock	0.0680	[0.0579	0.0779]	0.05	∞	IG
σ_{ξ^x}	std. dev. of investment efficiency shock	0.0248	[0.0117	0.0383]	0.05	∞	IG
σ_{ξ^h}	std. dev. of labour supply shock	0.0381	[0.0121	0.0691]	0.05	∞	IG
σ_p	std. dev. of probability of tax audit	0.0381	[0.0121	0.0691]	0.05	∞	IG

Note: official economy (OE), underground economy (UE), inverse-gamma (IG).

Table 2 Prior and posterior distributions

At first, we estimate a log-linearised form of the DSGE model, described in the previous section, using Bayesian techniques. Since the posterior probability distribution for this model cannot be computed analytically, we need to use approximation by applying Markov Chain Monte Carlo techniques. To be more precise, we apply the Random Walk Chain Metropolis–Hastings (RWMH) algorithm using 200000 replications and 100000 burned-in samples. The prior and posterior distributions of the estimated parameters are presented in the Table 2. It should be noted, that the log-linearised specification of the model contains the steady-states of endogenous variables that depend on the model parameters. We have iterated the process of estimation to the point where the estimated mean values of the parameters are consistent with the steady-state values of the variables used within the estimation procedure.

Using the mean values of estimated parameters, we start to solve the non-linear form of our model to get the steady-state values of all endogenous variables. We refer to this model as the benchmark model. Using the data, we have calibrated the steady-state values of the personal income tax rate at 0.172, corporate tax rate at 0.22, social security tax rate at 0.36, and the probability to be investigated (probability of tax audit) at 0.012. All values reflect the long-term averages. Furthermore, we perform a set of simulations that allow us to evaluate the effects of changes in the corporate tax rate, personal income tax rate, social security tax rate, and in the probability of tax control. The Laffer curves describe the simulated responses of the steady-state values of government revenues and output to the changes in the model tax rates. In our simulations, the government revenues are expressed as a ratio to the total government revenues of the benchmark model, and the total output and the output in the official economy are expressed as a ratio to the total output of the benchmark model. The size of the shadow economy is measured as a share of the total (official and unofficial) output of the economy.

4 Laffer curves in Slovakia

In this section, we discuss the impacts of alternative tax adjustments on the government tax revenues and the size of the shadow economy in Slovakia. Parameter estimates of the benchmark model imply the average size of the shadow economy to be 7.2% of the total output. This value is lower than the estimates of 12% provided by Quintano et al. [5]. It may indicate the decrease in the size of shadow economy in Slovakia since 2010. Our model suggests that 5.8% of the labour force is working in the underground economy. This result is slightly higher when compared with the estimated value of 3% presented by Schneider [6] from the year 2006. Above mentioned results may indicate the overall rise in the involvement of the Slovak workers in less productive economic activities within the shadow economy (on the other hand, the differences in the years of estimates provided by both studies may matter as well).

Figure 1 depicts the changes in steady-state values of the output in the official economy, total tax revenues, and the size of the shadow economy due to changes in the corresponding tax rates. The Laffer curves of personal income taxation and corporate taxation have expected shape. The peak of both curves may be found at the tax rate of 40% (slightly higher peak value is related to the corporate tax rate). The increase of personal income tax rate leads to a response of the size of the shadow economy, especially when crossing the peak values. The workers are thus more sensitive to the changes in income taxation. An interesting feature of the model is the Laffer curve of social security taxation. In this case, we are not able to observe the decrease of the total government revenues as the result of increasing tax rate. The reason is that this tax is modelled as tax-deductible cost and it does not influence the incentives of firms to produce the goods in the underground sector. It is thus hard for the workers to find the jobs in the shadow economy for a higher wage. The simulated increase of the total output is induced by the increased level of capital and by the increased labour supply willing to compensate for the decreased real income by increasing its labour effort.

5 Conclusion

In this paper, we estimated the small DSGE model consisting of both formal and informal sectors of the economy. Our estimates, based on the data of the Slovak economy, imply the size of the shadow economy to be 7.2% of the total output where 5.8% of the labour force is working in the underground economy. These results indicate the overall decrease of the size of the underground economy since 2010 after comparing our estimates with those from the previous studies. The Laffer curves of personal income taxation and corporate taxation have expected shape. The peak of both curves may be found at the tax rate of 40% that is higher than the actual tax rates. An interesting feature of the estimated model is a small sensitivity of the size of the shadow economy to the changes in social security taxation. The increase of the social security tax rates causes only small growth in the share of the shadow economy without negative effects on total government tax revenues.

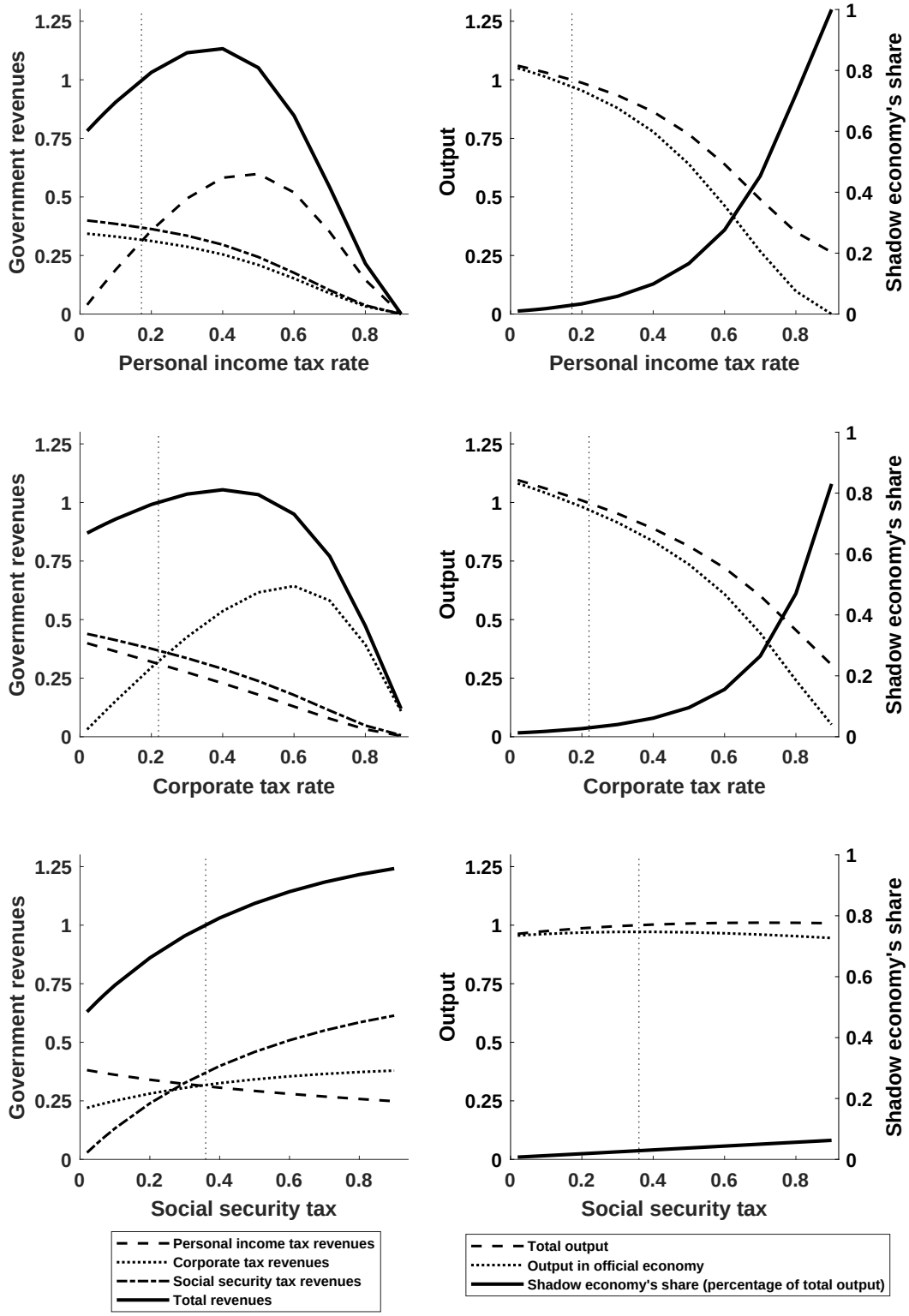


Figure 1 Simulated Laffer curves in Slovakia

Note: Government revenues are expressed as a ratio to the total government revenues of the benchmark model. Total output and the output in the official economy are expressed as a ratio to the total output of the benchmark model. The size of the shadow economy is measured as a share of the total (official and unofficial) output of the economy. The vertical dotted lines indicate the corresponding tax rates in the benchmark model.

Acknowledgements

This research was funded by funding for specific research at the Masaryk University, Faculty of Economics and Administration, project MUNI/A/0972/2018.

References

[1] Feld, L. P. & Schneider, F. (2010). Survey on the Shadow Economy and Undeclared Earnings in OECD Countries. *German Economic Review*. 11(2), 109-–149.

[2] Kirchgassner, G. (2017). On Estimating the Size of the Shadow Economy. *German Economic Review*. 18(1), 99–111.

[3] OECD, (2002). *Survey on the Shadow Economy and Undeclared Earnings in OECD Countries*.

[4] Orsi, R. & Raggi, D. & Turino, F. (2014). Size, Trend, and Policy Implications of the Underground Economy. *Review of Economic Dynamics*. 17(3), 417-–436.

[5] Quintano, C. & Mazzochi, P. (2010). Some Alternative Estimates of Underground Economies in 12 New EU Member States. *Iternational Economic Journal*. 24(4), 611–628.

[6] Schneider, F. (2011). *The Shadow Economy and Shadow Economy Labour Force: What Do We (Not) Know?*. IZA Discussion Paper Series, No. 5769.

[7] Schneider, F. & Buhn, A. (2016). *Estimating the Size of the Shadow Economy: Methods, Problems and Open Questions*. IZA Discussion Paper Series, No. 9820.

[8] Solis-Garcia, M. & Xie, Y. (2018). Measuring the size of the shadow economy using a dynamic general equilibrium model with trends. *Journal of Macroeconomics*. 56, 258–275.

[9] Trabandt, M. & Uhlig, H. (2011). The Laffer curve revisited. *Journal of Monetary Economics, Elsevier*. 58(4), 305–327.

Rockets and Feathers Effect in Slovak Gasoline and Diesel Markets

Karol Szomolányi[1], Martin Lukáčik[2], Adriana Lukáčiková[3]

Abstract. Numerous studies dealing with the transmission of crude oil prices to retail gasoline prices indicate that retail gasoline prices respond more quickly when crude oil price rises rather than when it decreases. This asymmetric retail gasoline price adjustment is called a ”rockets and feathers” effect. An explanation of the effect can be addressed to the theory of strategic interactions between firm and its consumers. According to the theory, the adjustment costs of the firm are expressed by the linex form.

In the paper we derive an econometric specification of a gasoline and diesel price reaction function from the linex adjustment cost function. We analyze the retail gasoline and diesel prices published by Statistical Office of the Slovak Republic during the period since 2009 till 2019. Estimating the specification, with the generalized methods of moments and weekly data, we do not reject the rockets and feathers effect in Slovak gasoline and diesel market. The weekly average Slovak retail price bias is about 0.011 eurocents per each gasoline liter sold and about 0.015 eurocents per each diesel liter sold.

Keywords: rockets and feathers, retail gasoline and diesel prices, linex adjustment costs.

JEL Classification: C26, C51, E32
AMS Classification: 62M10**,** 91B84

1 Introduction

Numerous studies dealing with the transmission of crude oil prices to retail gasoline prices indicate that retail gasoline prices respond more quickly when crude oil price rises rather than when it decreases; e.g. [13], [5], [6], [9], [7], [14], [15]. Bacon [1] called this asymmetric retail gasoline price adjustment as ”rockets and feathers” effect. His study was followed by paper of Borenstein et al. [2] who provide strong evidence of asymmetry in the US market between 1986 and 1992 in different stages of the production and distribution of gasoline.

Douglas and Herrera [4] provide an explanation of the rockets and feathers effect. According to authors, asymmetric retail gasoline price reactions on the crude oil price changes can be addressed to the theory of strategic interactions between firm and its consumers; e.g. [11]. Assume that demand for gasoline and diesel has been steadily increasing. Next assume that a firm behaves altruistically according to the theory of strategic interactions between firm and its consumers. It means that it does not increase its price if demand increases, but – on the other hand – it increases price if crude oil price increases. Then we predict that after crude oil price rises, an altruistic firm will raise retail gasoline price softly higher than a selfish firm would. It is because an altruistic firm through fear of angering a minority of customers will not react directly when demand changes, while a selfish firm can react freely. One way, how to adjust prices to the demand increase for an altruistic firm, is rising gasoline price vehemently after crude oil price rises and lowering gasoline price slightly after crude oil price decreases.

The commonly used methods in given empirical studies are error correction models (ECM) and vector autoregressive models (VAR); e.g. [13], [5], [6], [7]. In the paper, we provide an alternative empirical approach based on linex adjustment costs formulation. Using similar approach Pfann and Palm [12] provided an example of asymmetry approach in a labour market. According to the approach the firing costs of manufacturing exceed the hiring costs. Our ideas are inspired by empirical studies analysing the U.S. and EMU monetary policy asymmetries provided by Surico [16], [17] and [18]. He responded on questions, whether central bankers weight

[1] University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, karol.szomolanyi@euba.sk.
[2] University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, martin.lukacik@euba.sk.
[3] University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, adriana.lukacikova@euba.sk.

differently positive and negative deviations of inflation, output and the interest rate from their reference values. This central banks' behaviour corresponds to the time consistency of monetary policy theory, e.g. [8] and [3].

The character of theory of strategic interactions between firm and its consumers [11] is similar to the character of asymmetry theory in a labour market [12] or time consistency of monetary policy theory [8], [3]. Because of the increasing gasoline and diesel demand, the adjustment costs of gasoline and diesel seller will be lower after the crude oil price rises and they will be higher, after the crude oil price lowers.

The aim of the paper is to derive an econometric specification of a gasoline and diesel price reaction function. Estimating the specification we will not reject the rockets and feathers effect in Slovak gasoline and diesel market. We analyse the retail gasoline and diesel prices published by Statistical Office of the Slovak Republic. The brief literature review is in the Introduction. The derivation of the specification is in the second section. Data and methodology is presented in the third section. The results of the paper are in the fourth section and the fith section concludes the paper.

2 Model

Consider that gasoline and diesel seller reacts asymmetrically on changes in crude oil price. According to the theory of strategic interactions between firm and its consumers [11], his adjustment costs will be lower after the crude oil price rises and they will be higher, after the crude oil price lowers. Therefore, after the fashion of Surico [16], [17] and [18] we consider the adjustment costs function F is in the linex form:

$$F\left[p_t, E_{t-1}\left(c_t\right)\right] = \frac{-\gamma\left[p_t - kE_{t-1}\left(c_t\right)\right] + e^{\gamma\left[p_t - kE_{t-1}\left(c_t\right)\right]} - 1}{\gamma^2} \tag{1}$$

where p_t is retail gasoline or diesel price, c_t is crude oil price, k is technology coefficient and γ is an asymmetry coefficient. A negative value of the coefficient γ implies that a negative value of the difference $p_t - kE_{t-1}(c_t)$ causes higher costs to the price-maker than it would if γ were positive. The linex specification nests the quadratic form as a special case, so that applying l'Hôpital's rule twice when γ tends to zero results in a reducing in the loss function (1) to the following symmetric parameterization:

$$\lim_{\gamma \to 0}\left\{F\left[p_t, E_{t-1}\left(c_t\right)\right]\right\} = \frac{1}{2}\left[p_t - kE_{t-1}\left(c_t\right)\right]^2 \tag{2}$$

The fuel price-maker chooses p_t in order to minimize the cost function (1). The first-order condition with respect to p_t is in the form:

$$\frac{-1 + e^{\gamma\left[p_t - kE_{t-1}\left(c_t\right)\right]}}{\gamma} = 0 \tag{3}$$

Condition (3) is a general description of the reaction function of the fuel price-maker. Performing the second-order Taylor expansion of the exponential terms in (3), we gain:

$$p_t - kE_{t-1}\left(c_t\right) + \frac{\gamma}{2}\left[p_t - kE_{t-1}\left(c_t\right)\right]^2 + v_t = 0 \tag{4}$$

The remainder of the approximation is v_t and it contains terms of the third or higher orders of the expansion.

We solve equation (4) for p_t and, prior to generalised method of moment's estimation (GMM) of the short-run relation, we replace expected values with actual values and we take the first differences of the relation. In practise we estimate the following nonlinear specification:

$$\Delta p_t = k\Delta c_t - \frac{1}{2}\gamma\Delta\left[\left(p_t - kc_t\right)^2\right] + u_t \tag{5}$$

From (3), we can also express the average gasoline or diesel price bias from the one-unit oil price increase caused by the rockets and feathers effect, $\gamma < 0$. Assuming that oil shocks Δc_t is a normally distributed process with zero mean and variance σ^2, taking the first differences, expected values and logarithms of (3) and after rearranging terms we gain the price bias in the form:

$$E\left(\Delta p_t\right) = -\frac{k^2\gamma}{2}\sigma^2 \tag{6}$$

3 Data and Methodology

Data of retail gasoline and diesel prices on Slovak market are gathered from the Statistical office of the Slovak Republic[4]. The spot prices for crude oil are gathered from the U.S. Energy Information Administration[5] – the agency responsible for collecting, analysing, and disseminating energy information. Since we have only the weekly retail gasoline and diesel prices data, we could only use the weekly Europe Brent Spot Price FOB Dollars per Barrel for the analysis.

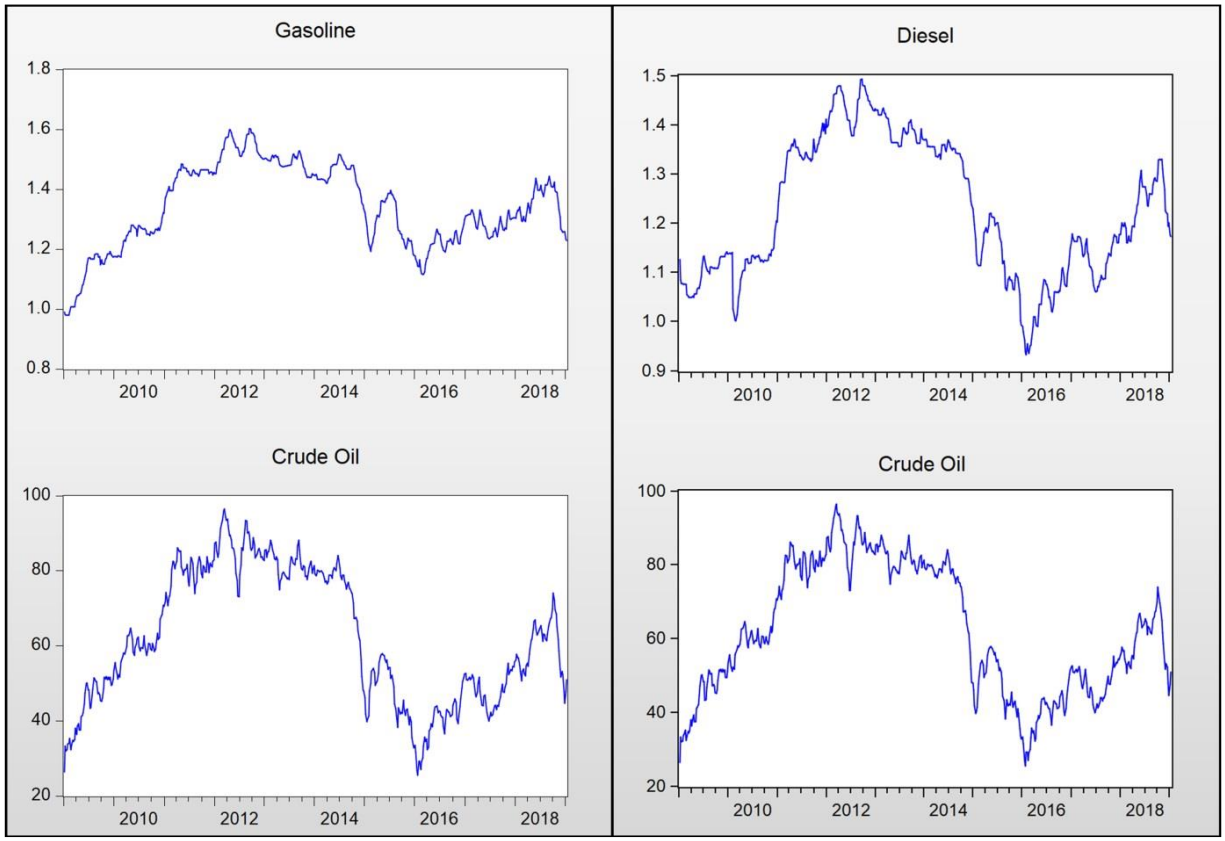


Figure 1 The weekly retail gasoline and diesel prices in euros (on the top, from the left) compared with the weekly crude oil prices in euros (on the bottom)

Liu et al [7] notice that taxes and levies make up a significant proportion of retail fuel prices and any changes in government taxes and levies can therefore have a significant impact on retail diesel and petrol prices. During the analysis period, there was no significant change in the consumption taxes.

The weekly retail gasoline and diesel prices data are in euros, so we needed to recalculate the crude oil prices from dollars to euros. We converted the daily oil prices in dollars by the euro exchange rate in dollars and then aggregated them into weekly averages. All data are with range since the first week of 2009 till the second week of 2019, 524 observations.

The orthogonality condition implied by the rational expectation hypothesis makes the general method of moments (GMM) a natural candidate to estimate equation (5) for both retail gasoline prices and diesel prices. The standard errors are computed with the procedure of Newey and West [10] in all estimates.

The one-period lags of the first differences of retail gasoline prices and crude oil prices were used as instruments in gasoline equation (5); the one-period lags of the first differences of retail diesel prices and crude oil prices were used as instruments in diesel equation (5), i.e. $\Delta p_{t-1}, \Delta c_{t-1}$ in both equations. By the orthogonality test of instruments, two model versions were estimated for all cases; one with an instrument and second without him. The difference in the corresponding J statistics is subject to χ^2 statistics with one degree of freedom (one instru-

[4] https://slovak.statistics.sk
[5] https://www.eia.gov/

ment). The corresponding results of the orthogonality test are in the Table 1. We do not reject the validity of instruments in all cases.

Equation	*instrument*	*Difference in J*
Gasoline	Δp_{t-1}	0.02
	Δc_{t-1}	0.40
Diesel	Δp_{t-1}	0.04
	Δc_{t-1}	0.04

Table 1 Orthogonality test of instruments

In our study, we made several other experiments. We are not going to present them in details, as these experiments do not unfix our results. Due to the robustness, the different methods of estimates of (5) (ordinary least square, two-stage least square, one-period-forward-looking generalized method of moments and system generalized method of moments) were performed. The combinations of one-period lags of the first differences of retail gasoline prices, retail diesel prices and crude oil prices were used as instruments. In the case of forward-looking GMM equations, not-lagged or not-leaded first differences of crude oil price data series is used as one of instruments. By the orthogonality test of instruments we state that all used instruments are valid. The results of different methods' estimates do not significantly differ from our results.

Expanding the quadratic term on the right side of (5), we obtain the linear econometric specification. The linear specification allows us to use Hausman test of endogeneity of regressors. We estimated the linear specification of gasoline prices with GMM method. The constant and the first differences of one-period lagged gasoline and oil prices, Δc_{t-1}, Δp_{t-1}, their product $\Delta(p_{t-1}c_{t-1})$ and their second powers, $\Delta({p_{t-1}}^2)$, $\Delta({c_{t-1}}^2)$ where used as instruments for the linear specification estimate. Using the nonlinear coefficient restriction tests we confirmed that the linear specification fits to the linear specification (5). Performing the proper test, we confirmed the endogeneity of regressors (stating that instruments choice in the estimate of nonlinear specification (5) mentioned above was correct).

4 Results

The estimates of (5) for gasoline and diesel prices are in the Table 2. The corresponding standard errors computed with the Newey and West [10] procedure are in parentheses. All estimates are statistically significant at the 1% level. In the fourth column of the table, the corresponding *J* statistics are. We also estimated price biases (6). Multiplying them with the mean of the oil price's first positive differences we estimated the weekly biases per a liter of gasoline or diesel. In the last column of the table, the weekly biases in eurocents per liter of sold gasoline and diesel are shown.

Equation	k	γ	J	Weekly Bias
Gasoline	0.008^{***}	-1.128^{***}	0.403	0.0106
	(0.002)	(0.113)		
Diesel	0.008^{***}	-1.404^{***}	0.043	0.0149
	(0.002)	(0.172)		

Table 2 GMM Estimates of (5)

Comparing the results for gasoline and diesel, we state that asymmetry coefficients are higher in diesel equations than in gasoline. Also estimated price biases are higher for diesel than for gasoline in all cases. It means that the rockets and feathers effect is higher in the Slovak diesel market.

Similarly as studies of different fuel markets [13], [5], [6], [9], [7], [14], [15], [1] and [2], our study confirms the rockets and feathers hypothesis in the Slovak fuel market as well. We used a different method as [13], [5], [6], [7], however our results are the same. Our methodology and specification are derived from the theory of strategic interactions between firm and its consumers (e.g. [4] and [11]) that can meaningfully explain the rockets and feathers hypothesis.

5 Conclusion

By estimating a reaction function based on linex adjustment costs formulation we confirmed the rockets and feathers effect in the Slovak gasoline and diesel market. Slovak retail gasoline and diesel prices respond more quickly when crude oil price rises rather than when it decreases. This effect is higher for diesel than for gasoline.

The linex adjustment costs formulation corresponds to theory of strategic interactions between firm and its consumers. This theory is one possible explanation off the New Keynesian price stickiness' theory of business cycles. The rockets and feathers effect at Slovak gasoline and diesel market supports the New Keynesian effects in business cycles. This result therefore may be useful for further studies of Slovak economic policy.

Acknowledgements

The Grant Agency of Slovak Republic - VEGA, supports this paper by grant no. 1/0294/18, "Short-run and long-run dynamic analysis of economic development of European post-communist countries and regions".

References

[1] Bacon, R. (1991). Rockets and Feathers: The Asymmetric Speed of Adjustment of UK Retail Gasoline Prices to Cost Changes. *Energy Economics, 13,* 211–218.
[2] Borenstein, S. & Cameron, A. & Gilbert, R. (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *Quarterly Journal of Economics*, *112*, 305–339.
[3] Calvo, G. (1978). On the Time Consistency of Optimal Policy in a Monetary Economy. *Econometrica, 46,* 1411–1428.
[4] Douglas, C. & Herrera, A. (2010). Why Are Gasoline Prices Sticky? A Test of Alternative Models of Price Adjustment. *Journal of Applied Econometrics, 25*, 549–579.
[5] Grasso, M. & Mannera, M. (2007). Asymmetric Error Correction Models for the Oil–Gasoline Price Relationship. *Energy Economics, 35*, 155–157.
[6] Honarvar, A. (2009). Asymmetry in Retail Gasoline and Crude Oil Price Movements in the United States: an Application of Hidden Cointegration Technique. *Energy Policy, 31,* 395–402.
[7] Liu, M., Margaritis, D. & Tourani-Rad, A. (2010). Is there an Asymmetry in the Response of Diesel and Petrol Prices to Crude Oil Price Changes? Evidence from New Zealand. *Energy Economics*, *32*, 926–932.
[8] Kydland, F. & Prescott, E. (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy*, *85*, 473–492.
[9] Meyler, A. (2009). The Pass-Through of Oil Prices into Euro Area Consumer Liquid Fuel Prices in an Environment of High and Volatile Oil Prices. *Energy Economics*, *31*, 867–881.
[10] Newey, W. & West, K. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*, 703–708.
[11] Okun A. (1981). *Prices and Quantities: A Macroeconomic Analysis*. Washington, DC: Brookings Institution.
[12] Pfann, G. & Palm, F. (1993). Asymmetric Adjustment Costs in Non-linear Labour Models for the Netherlands and U.K. Manufacturing Sectors. *Review of Economic Studies*, *60*, 397–412.
[13] Radchenko, S. (2005). Oil Price Volatility and the Asymmetric Response of Gasoline Prices to Oil Price Increases and Decreases. *Energy Economics*, *27*, 708–730.
[14] Rahman, S. (2016). Another Perspective on Gasoline Price Responses to Crude Oil Price Changes. *Energy Economics*, *55*, 10–18.
[15] Sun, Y. & Zhang, X., Hong, Y. & Wang, S. (2018). Asymmetric pass-through of oil prices to gasoline prices with interval time series modelling. *Energy Economics, 78,* 165–173.
[16] Surico, P. (2007). The Fed's Monetary Policy Rule and US inflation: the Case of Asymmetric Preferences. *Journal of Economics Dynamics and Control*, *31*, 305–324.
[17] Surico, P. (2007). The Monetary Policy of the European Central Bank. *Scandinavian Journal of Economics, 109*, 115–135.
[18] Surico, P. (2008): Measuring the Time Inconsistency of US Monetary Policy. *Economica*, *75*, 22–38.

Semantic Model of Project Management in Corporate practice

Bartoška Jan[1], Jedlanová Tereza[2], Rydval Jan[3]

Abstract. The paper proposes the use of semantic networks and analytical network analysis (ANP) methods for quantification of the "soft" structure of a project in corporate organization. The semantic project networks are derived based on the organization structure and on the life cycle of projects. Their subsequent quantification using the ANP method creates the basis for analysis of the project roles and analysis of their individual relationship to organization units. Although the project roles during the management of projects are the key to the success, an approach to the quantification and analysis of their impact and influence on project organization structure has not been introduced yet. Semantic networks can be used to manage projects to illustrate and quantify links between internal and external objects of the project environment - project goals, project outputs, project documents, etc. The paper suggests a new approach for identifying and quantifying project roles in the project organization management. The showed semantic model and its quantification can be used to construct a multi-criteria model for decision support in commercial practice. The paper contains a case study of application of semantic model of project management in commercial organization from bank sector.

Keywords: Project Management, Corporate Organization, Semantic Networks, Stakeholder Management, Analytic Network Process, Human Resource Project Management, Multi-criteria Decision Making.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The development of organizations, project-oriented in particular, is currently at a glance - the transition from classic waterfall management to agile management is currently evident in most large corporations in telecommunications and banking services, not only in the Czech Republic. The significance of project roles, i.e. individual job positions inside and outside the project team, is growing in importance - the need for effective changes in internal project management methodologies and procedures and project teams is also growing in importance.

Project management is a collection of many professional disciplines and skills when, besides traditional quantitative approach, qualitative approach which is the content of international standards is gaining on importance. International standard or methodology of project management PMBOK® Guide [6] or PRINCE2 [1] describes a common practice to set organizational, instrumentation, process and knowledge management aspects of a project, giving rise to the so-called "soft" structure of the project, i.e. a partially ordered set of documents, roles, workflows and tools with variable influence on internal and external environment of the project. Although the prioritization of project roles in communication and project documentation in project management is a key to the success of the project, a unique approach to quantifying the "soft" structure of the project has not yet been introduced. "Soft" structures of the project are usually just displayed in the form of knowledge maps or semantic networks. For potential quantification it is appropriate to focus particularly on the use of semantic networks.

Semantic networks are conventionally utilized in ICT, especially in developing and maintaining software and web applications; e.g., Zhu and Li [10] focus on developing a semantic approach in IT for the resolution of an application and contextual level. Whereas Williams [9] uses a semantic network for the display and analysis of success factors in the structure of mutual expectations and limitations of the project stakeholders. While El-Gohary, Osman and El-Diraby [4] for example indicate that insufficient or inadequate stakeholders' involvement in the management of projects, programs and portfolios is the main reason for failure, they further propose the

[1]Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, jedlanova@pef.czu.cz.
[3] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, rydval@pef.czu.cz.

semantic model as a tool to visualize and manage relationships and knowledge towards stakeholders in a multi-project environment. Following the example of their use in IT, it is therefore appropriate to continue to develop semantic models to meet the needs of project management, i.e. a management structure of roles, documentation, and knowledge constraints in a project.

The aim of the article is to demonstrate and interpret the results and benefits of the use of the semantic model of project management in the commercial sphere, namely in the banking sector of the selected organization. The contribution of the article is to differentiate the subjective attitudes of respondents, i.e. project roles, to superior authorities and departments with the help of quantifying the semantic network using the ANP model.

2 Materials and methods

2.1 International standard and methodology of the project management

The international standard and project management methodology "Guide to Project Management Points of Knowledge (PMBOK®Guide)" [6] and "Managing Successful Projects with PRINCE2" [1] form a methodological basis for project-oriented organizations in today's practice. In particular, this minimum recommended practice has been applied in the banking sector.

The *PMBOK® Guide* [6] is process-oriented and in its knowledge areas it presents the set of qualitative (e.g. team management, team motivation, etc.) or quantitative approaches (e.g. CPM, EVM, etc.) necessary for project management. Knowledge areas of the *PMBOK®Guide* standard represent the substantive content of the project lifecycle processes. The standard thus dissolves project management into atomic transformations where the entry is secured, necessary tasks (expert, professional, managerial, etc.) are performed and output is created. Outputs from one process become inputs of the second.

The topics of *PRINCE2* methodology are as follows [1]: Business Case; Organization; Quality; Plans; Risk; Change; Progress. The seven processes which *PRINCE2* methodology provides describe the life cycle of a project from the position of roles and responsibilities. All seven *PRINCE2* processes can be continuously connected. *PRINCE2* methodology processes are based on the interaction of project roles and stakeholders in the organizational structure of the project. Project roles and roles of stakeholders in *PRINCE2* are as follows: Steering Committee; Sponsor; Senior User; Senior Supplier; Project Manager; Team Manager; Change Authority; Project Support; Project Assurance.

2.2 Semantic model and Analytic network process (ANP)

The project management structure in the organization can be described in the form of a semantic network [2], both in the commercial and non-commercial spheres, as described in [3] or [5]. It is particularly useful to distinguish project roles, project documentation, project constraints, knowledge management areas and interrelationships between elements or groups of elements, see Figure 1 as states [3].

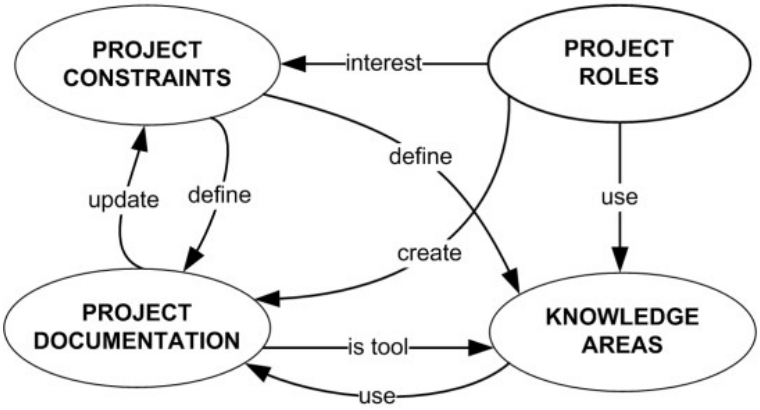


Figure 1 Semantic Model of Project Structure in the ANP model [3].

The project or project management semantic network defined in the organization can be simplified (without link interpretation) into the Analytic Network Process (ANP) model and then used to analyze the organization's environment, see Figure 2 as reported in [3].

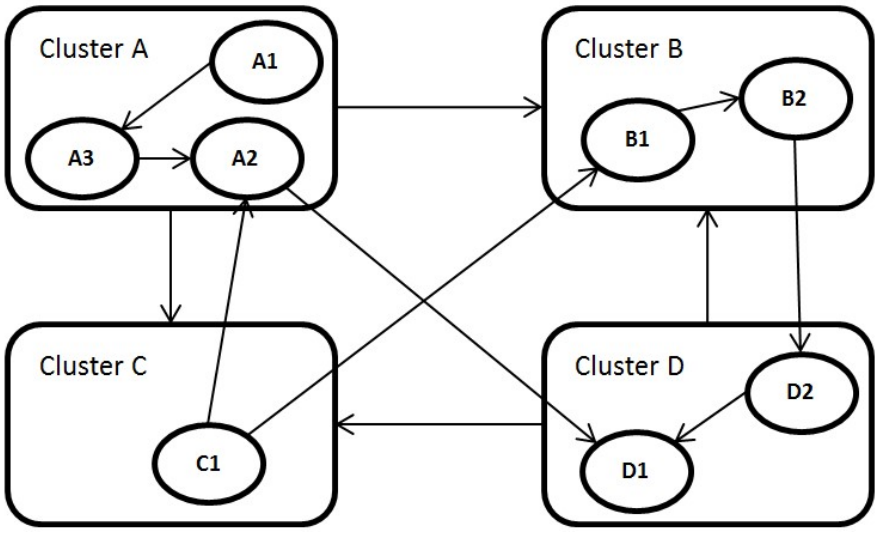


Figure 2 The **s**tructure of elements and clusters in ANP model [3].

The benefit of the ANP method is the ability to express different preferences of links between elements and clusters. To express preferences the method of pair comparison is used. Preferences always occur precisely in assessing the importance of the two elements in terms of the element which refers to them - there rises a question "which of the elements is more important, and by how much". The resulting values for the sub-clusters are then combined into a super matrix where the normalization of columns is performed [8], [7]:

$$W = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_N \end{array} \\ \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix} \end{array} \tag{1}$$

Where each block of the super matrix consists of:

$$W_{ij} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \tag{2}$$

Under condition:

$$\sum_{i}^{n} w_{ij} = 1, j \in \langle 1, n \rangle \tag{3}$$

For a weighted super matrix (1) for which the relation (3) is valid, a calculation to obtain limit weights of the elements can be performed. The calculation is performed by squaring the weighted super matrix to a sufficiently large number. Since the super matrix has *N*x*N* size, the squaring is always feasible in a trivial manner (matrix multiplication). The result is the approximation of the weighted matrix to the limit matrix. Limit scales can be found in any column of the super matrix. The limit weight of each element expresses the strength of the effect on the overall structure of elements, i.e. it answers the question of how strongly an element affects the other elements [8], [7].

2.3 Case study: Organization project structure in bank sector

Currently, a multi-level project organizational structure has pushed through in the current project management practice in the corporate environment with the prevailing standard of "Guide to Project Management Body (PMBOK®Guide)" [6] and the methodology "Managing Successful Projects with PRINCE2" [1], which includes both project roles (PgM, PM), Project Leadership (PSC) and senior corporate units (PPM, PMO, BoD, EAB, PROB):

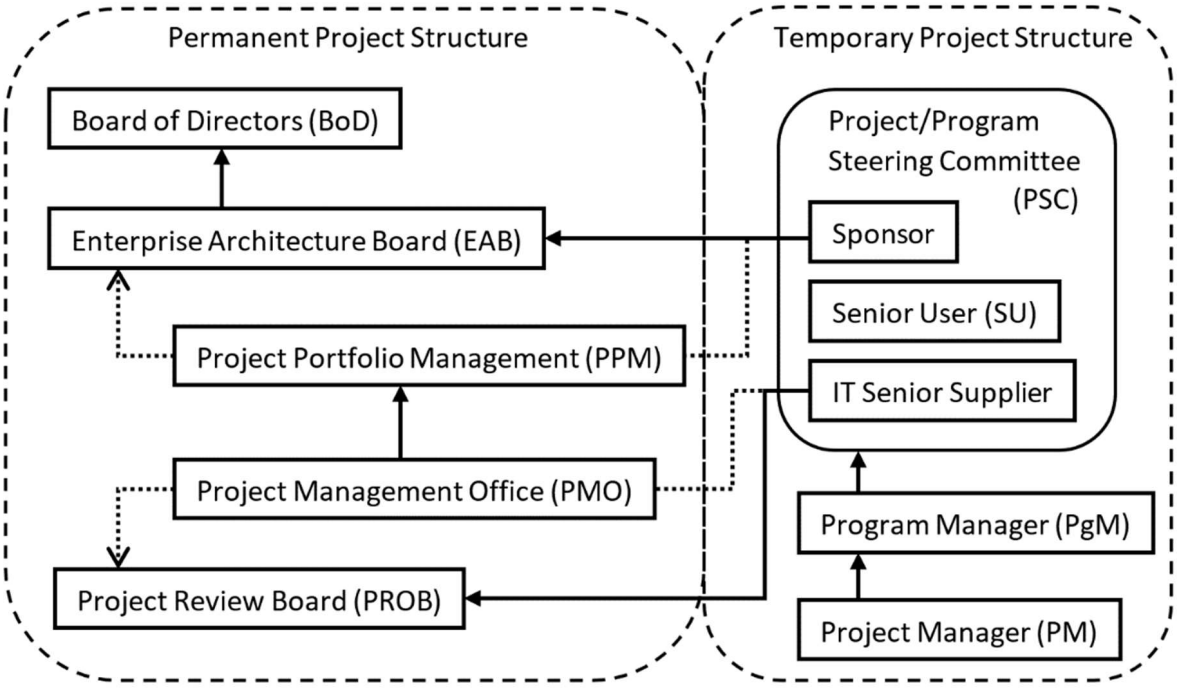


Figure 3 The organization project structure in bank sector.

Figure 3 shows the interaction between project roles (PgM, PM), the project's superior body (PSC), and corporate units (BoD, EAB, PPM, PMO, PROB – without participation of the project roles). Project roles (PgM, PM) report on the progress and success of individual projects, while the superior PSC decides on changes and key life situations of individual projects (start, close, etc.). The project roles of PgM and PM and the parent PSC are temporary structures. The number of managed projects does not affect the influence of the project role in the organization project structure. The BoD, EAB, PPM, PMO and PROB corporate units are permanent units that provide strategic project management in the organization. The organization project structure will be different for each the corporate organization. Typically, the internal project management methodology of the organization defines and interacts with the interaction between the roles, the authority and the departments (units) in the context and in accordance with the international standard "Guide to the Project Management of Knowledge (PM-BOK®Guide)" [6] or the "Managing Successful Projects with PRINCE2" methodology [1].

3 Results and discussion

The research in the chosen bank organization took place from 2016 to 2018. Within the research, a basic semantic model of project management was created as stated in [2] and further described and interpreted in [3] - the model includes a complete network of project roles, departments, project documentation, project restrictions, etc. The model was first quantified, i.e. the limit weights of the elements were determined, without project role preferences (Neutral model without preferences).

This model was then repeatedly discussed with selected banking organization staff, with subjective preference recording, and always stored and quantified in the copy - thereby obtaining different limit weights for individual elements, incorporating individual respondent preferences. Two program managers (PgM 1, PgM 2) and four project managers (PM 1, PM 2, PM 3, PM 4) were selected among respondents. For the purpose of this paper, only the limiting weights of superior departments and bodies (EAB, PMO, PPM, PROB, PSC) for the basic model and models with individual respondent preferences were selected. The article presents the most interesting part of the research - subjective differentiation of project roles.

The semantic model quantification of the project structure organization can be performed by the ANP method. The advantage of this method is the possibility of bias preferences among elements – between project roles and organization units. For example, with the help of the Saaty scale [7], the addressed project roles can differentiate their different attitude toward the superior authorities (units from Permanent project structure). Creating the calculation of the ANP model can be performed e.g. in a software tool Super Decisions Software 2.1 (*http://www.superdecisions.com/*). By calculating Calculus Type, a super matrix with limit weights can be obtained.

		Neutral model without preferences	**Project Roles** from **Temporary Project Structure**					
			PgM 1	*PgM 2*	*PM 1*	*PM 2*	*PM 3*	*PM 4*
Superior Units from **Permanent Project Structure**	*EAB*	0.13018	0.14139	0.20117	0.11991	0.12378	0.13017	0.19361
	PMO	0.14465	0.15976	0.10787	0.18515	0.10540	0.14468	0.14541
	PPM	0.29504	0.23215	0.26135	0.24981	0.20852	0.29504	0.38075
	PROB	0.14442	0.13597	0.18784	0.10184	0.24008	0.15439	0.02782
	PSC	0.28571	0.33073	0.24177	0.34329	0.32222	0.28571	0.25241

Table 1 Limit weights of project roles towards the superior units.

The values of limit weights for individual respondents, i.e. for the addressed Program Managers (PgM) and Project Managers (PM) differ from their superior bodies in the project organizational structure. Different values in Table 1 reveal different attitudes of individual respondents. The basic ANP model, which does not contain any individual preferences, expresses the expected significance of the given body (superior body) according to the internal project management methodology of the organization - i.e. the expected "level" according to the international standard PMBOK® Guide [6] or PRINCE2 methodology. [1]. In practice, however, in project management, when interacting with respondents, the significance of departments in the project structure changes or varies.

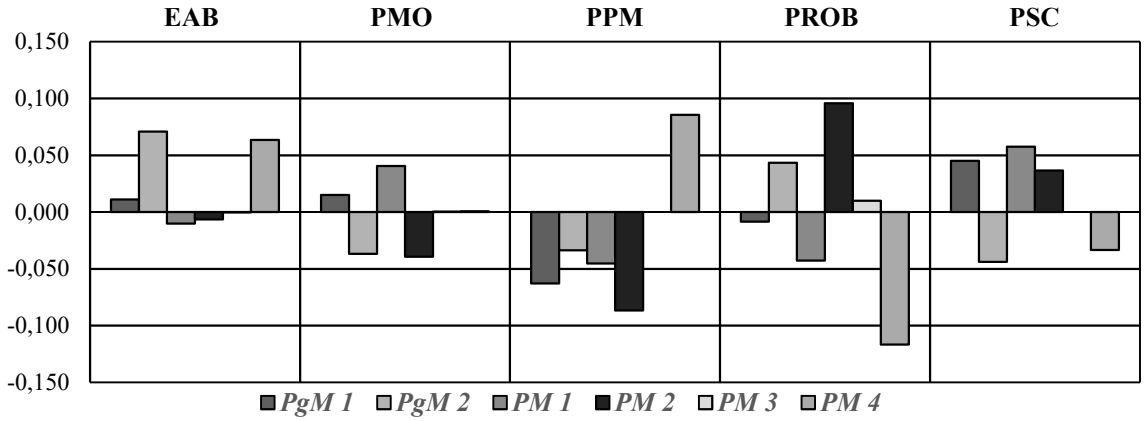


Figure 4 The differences of limit weights of the project roles towards the superior units.

The biggest differences among respondents from the base model were found against the Project Portfolio Management (PPM) and Project Review Board (PROB) as shown in Figure 4. The lowest differences or "biases" from the base model are seen against the Project Management Office (PMO), see figure 4. The limit weights of respondents express their professional and possibly personal attitudes - the significance of departments and superior bodies varies when individual respondents interact. Differences in the weights of individual respondents may be a topic for further development of the organization, changes in internal project management methodology or changes in organizational project structure. Responding variations in limit weights can also be seen as a characteristic of the "credibility" or "timeliness" of internal management methodology, i.e. methods, procedures and rules for managing and preparing projects in an organization.

4 Conclusion

The paper presents the use of semantic models of project management in the commercial sector, specifically in the banking sector. The presented results are derived from partial results of the authors' own research from 2016 to 2018. The results of the case study show that it is possible to quantify the individual attitude of the project roles (stakeholders) to the superior departments (bodies) in the project organizational structure. Furthermore, the discrepancies identified can serve to further analyse the situation in the organization and to improve project management approaches and processes.

The paper presents the results of quantification of semantic models of project management using the ANP model, using the subjective preference expression according to the Saaty scale. Reusing this process and quantifying it can lead to a meaningful analysis of the organization and a targeted development of internal project management methodologies and procedures.

Acknowledgements

Acknowledgements This research is supported by the grant No. 2019A0015 “Ověření a rozvoj sémantického modelu řízení projektů ” of the Internal Grant Agency of the University of Life Sciences Prague.

References

[1] Axelos. (2017) *Managing Successful Projects with PRINCE2 2017th Edition*. London: The Stationery Office. ISBN 978-0113315338.

[2] Bartoška, J. (2016). *A sematic model for the management of projects*. Habilitation work (thesis). DSE FEM CULS Prague.

[3] Bartoška, J. (2016). Semantic Model of Organizational Project Structure. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Liberec: Technical University. ISBN 978-80-7494-296-9.

[4] El-Gohary, N. M., Osman, H., El-Diraby, T.E. (2006) Stakeholder management for public private partnerships. *International Journal of Project Management*, *24*, 595-604.

[5] Jedlanová, T., Bartoška, J., Vyskočilová, K. (2018). Semantic Model of Management in Student Projects. In: *Proceedings of the 36th International Conference Mathematical Methods in Economics*. Prague: MatfyzPress. ISBN 978-80-7378-372-3.

[6] PMI. (2017) *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) — Sixth Edition and Agile Practice Guide*. Newtown Square (PA): Project Management Institute Inc. ISBN 978-1628253825.

[7] Saaty, T. L. (1996) *Decision Making with Dependence and Feedback: The Analytic Network Process*. Pittsburgh: RWS Publisher. ISBN 0-9620317-9-8.

[8] Saaty, T. L., Vargas, L. G. (2006) Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Application with Benefits, Opportunities, Costs and Risks. New York: Springer. ISBN 978-0-387-33987-0.

[9] Williams, T. (2015) Identifying Success Factors in Construction Projects: A Case study. *Project Management Journal*, *47*, 97-112.

[10] Zhu, Y., Li, X. (2007) Representations of semantic mappings: A step towards a dichotomy of application semantics and contextual semantics. *International Journal of Project Management*, *25*, 121-127.

Analysis of labour market development in the Czech Republic

Radmila Krkošková[1]

Abstract. This article deals with an analysis of the relationship between the unemployment rate and the number of vacancies in the labour market. This relationship can be characterized by the Beveridge curve. The article explores, whether there is a long-term or short-term relationship between variables. Furthermore, this relationship in the labour market is expanded with the impact of the rate of the unemployed earning unemployment benefit, the rate of unemployed graduates and the rate of unemployed women. The vector error correction model was used for this purpose to determine both long-term and short-term causal relationships. To create the resulting model, the econometric methodology was used, namely unit root tests, Granger causality for the determination of statistically significant relationships, information criteria and the Johansen cointegration test. The data used have the character of quarterly time series in the period from 2002Q1 to 2018Q4. EViews software version 9 was used for the calculations.

Keywords: ADF test of stationarity, Beveridge curve, cointegration test, Johansen test, labor market, VECM

JEL Classification: C22, C32, J64
AMS Classification: 62C07

1 Introduction

The level of unemployment, often expressed as the unemployment rate, and the number of vacancies recorded, expressed as the job vacancy rate, are monitored and evaluated in analyses of labour markets at the macroeconomic level, as well as the mutual relationship between these variables. The Beveridge curve can be used to analyse this relationship. This approach provides basic data on the labour market situation, as reported by Filer *et al.* [9]. Using a Beveridge curve, we can also identify the types of unemployment that occur. The model Blanchard & Diamond [5] is considered a basic theoretical study dealing with this relationship.

The core of this paper is the analysis of two hypotheses concerning the relationship between the unemployment rate and the rate of job vacancies in the Czech Republic: H1: There is a long-term relationship between the unemployment rate and the rate of job vacancies in the Czech Republic, H2: There is a short-term relationship between the unemployment rate and the rate of job vacancies in the Czech Republic. The positive relationship between the variables is assumed.

The labour market is affected by both external and internal conditions, as stated in the article from Doležalová [8]. The most important internal factors related to labour supply include: population literacy, age structure, duration of unemployment, mobility, number of the unemployed. Important factors on the side of the demand include: characteristics of job vacancies, number of vacancies. The mutual relationship between labour supply and demand is very important for the overall characteristics of the labour market. If economy development is fluctuating, it is very important how quickly and flexibly the labour market can adapt to new conditions.

2 Theoretical background and data

2.1 Theoretical background

As already mentioned, the reciprocal relationship between unemployment and vacancies can be characterized by a Beveridge curve. This topic has been addressed by a number of economists, such as Bleakley & Fuhrer [6] and

[1] School of Business Administration in Karviná, Silesian University in Opava, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, e-mail: krkoskova@opf.slu.cz

Bewley [4] in UK. The following text is based on the work of Hančlová *et al*. [13]. The description of individual variables is shown in Table 1.

The absolute increase in the number of employed can be approximated by the following relationship:

$$\Delta E - E_t - E_{t-1} \equiv H_t - Q_t \tag{1}$$

where H_t is the number of new employment contracts, Q_t the number of terminated employments.

The function H_t can be expressed as a homogeneous Cobb-Douglas production function

$$H_t = b_0 \cdot UN_{t-1}^{\alpha} \cdot V_{t-1}^{1-\alpha} \tag{2}$$

where b_0 is the efficiency of job search, the share of successful contacts to the total number of contacts in search of work for the period t; UN_t is the number of registered at the labor office; V_t is the number of vacancies. The number of terminated employment relationships is linearly dependent on the number of employees at the time t – 1 with the parameter c_0. The absolute change in the number of employees is reflected in the non-equilibrium model of the labor market:

$$\Delta E_t = b_0 \cdot UN_{t-1}^{\alpha} \cdot V_{t-1}^{1-\alpha} - c_0 \cdot E_{t-1} \tag{3}$$

and the relative increase of employed e_t is given $\quad e_t = b_0 \cdot une_{t-1}^{\alpha} \cdot ve_{t-1}^{1-\alpha} - c_0$, (4)

where $e_t = \dfrac{\Delta E_t}{E_{t-1}}$; $\quad une_{t-1} = \dfrac{U_{t-1}}{E_{t-1}}$; $\quad ve_{t-1} = \dfrac{V_{t-1}}{E_{t-1}}$.

The long-term model of the labor market, which is derived in logarithmic form, respects condition that the amount of labor and labor resources are constant, that is $e_t = 0$:

$$\ln(une_{t-1}) = -\frac{1-\alpha}{\alpha} \cdot \ln(ve_{t-1}) + \frac{1}{\alpha} \cdot \ln\left(\frac{c_0}{b_0}\right) \tag{5}$$

Anthony [3], in his work on this deterministic model, introduced a random component into equation (3)

$$\Delta E_t = b_0 \cdot UN_{t-1}^{\alpha} \cdot V_{t-1}^{1-\alpha} \cdot \left\{e^{\mu_{t-1}}\right\} - c_0 \cdot \left\{e^{\rho_{t-1}}\right\} \cdot E_{t-1} \tag{6}$$

respectively in logarithmic form:

$$\ln(une_t) = -\frac{1-\alpha}{\alpha} \cdot \ln(ve_t) + \frac{1}{\alpha} \cdot \ln\left(\frac{c_0}{b_0}\right) + \frac{1}{\alpha} \cdot \varepsilon_t , \tag{7}$$

where $\varepsilon_t = \rho_t - \mu_t$; $\quad \mu_t \approx N\left(0, \sigma_\mu^2\right)$; $\quad \rho_t \approx N\left(0, \sigma_\rho^2\right)$.

Equation (7) allows testing the existence of a long-term relationship between variables une_t and ve_t assuming these variables are non-stationary and are integrated in order one. The labor model of the two variables (7) assumes that the parameters b_0, c_0 are fixed, that is the flows between the number of applicants and the number of vacancies is exogenous.

In part of 3.2, this model will be expanded. The model will assume that the flows between the number of job seekers and the number of vacancies are endogenous and are determined by the development of the labor market structure. This structure will present the time series: P_t - the number of registered applicants earning unemployment benefit; A_t - the number of registered graduates among job seekers; Z_t - the number of registered unemployed women at employment offices.

The modified non-equivalence model of the labor market is based on the adjustment in the equation (6):

$$\Delta E_t = b_0 \cdot q_t^b \cdot UN_{t-1}^{\alpha} \cdot V_{t-1}^{1-\alpha} \cdot \left\{e^{\mu_{t-1}}\right\} - c_0 \cdot \left\{e^{\rho_{t-1}}\right\} \cdot E_{t-1} \tag{8}$$

and the expanded equilibrium labor market model for the three variables (trivial model) is given by:

$$\ln(une_t) = -\frac{1-\alpha}{\alpha} \cdot \ln(ve_t) - \frac{b}{\alpha} \cdot \ln(qe_t) + \frac{1}{\alpha} \cdot \ln\left(\frac{c_0}{b_0}\right) + \frac{1}{\alpha} \cdot \varepsilon_t \tag{9}$$

$\varepsilon_t = \rho_t - \mu_t$; $\quad qe_t = \dfrac{q_t}{E_t}$, where q_t is a time series expressing the structure of the labor market (P_t, A_t, Z_t).

A similar topic is dealt with in the following articles. The impacts of education investment on skilled–unskilled wage inequality and economic development in developing countries are described in the article Pan [16]. Al-Ubaydli & List [2] are studying the problems of the labor market. Contractionary effects of the current global crisis are described in the paper Gündüz & Kaya [11]. Sectoral employment multipliers and high regional unemployment differentials unequally transfer effects into the labor market. The purpose of this study is to explore the inter-sectoral sensitivities of regional labor markets in Turkey to changes in the final demand. The article Abe [1] studies regional variations in labor force behavior of women in Japan. Short-run and long-run linkages between employment growth, inflation and output growth: evidence from a large panel we can find in the paper Škare & Caporale [19]. This study examines the short- and long-run linkages between employment growth, inflation and output growth applying panel cointegration and causality tests to data for 119 countries over the period 1970–2010. We find evidence of positive Granger causality running from output growth to employment growth in the short run. Employment growth Granger causes output growth with a negative sign in the long run. Modelling of unemployment in the Czech Republic is described in the follow papers: Čabla & Malá[7], Hudcovský *et al.* [14] and Flek & Myslíková [10].

2.2 Data

Quarterly data for the period from 2002/Q1 to 2018/Q4 were used for the calculations. The Czech Statistical Office and Ministry of Labour and Social Affairs were the primary data source, MPSV [15]. The description of individual variables is shown in the Table 1 and Figure 1. The selection of variables was done according to Hančlová *et al.* [13] and Hančlová & Šimek [12].

Variable designation	Description of variable
UN	The number of registered jobseekers at employment offices
V	The number of job vacancies
P	The number of registered applicants earning unemployment benefit
A	The number of registered graduates among job seekers
Z	The number of registered unemployed women at employment offices
E	The number of employed

Table 1 Description of variables

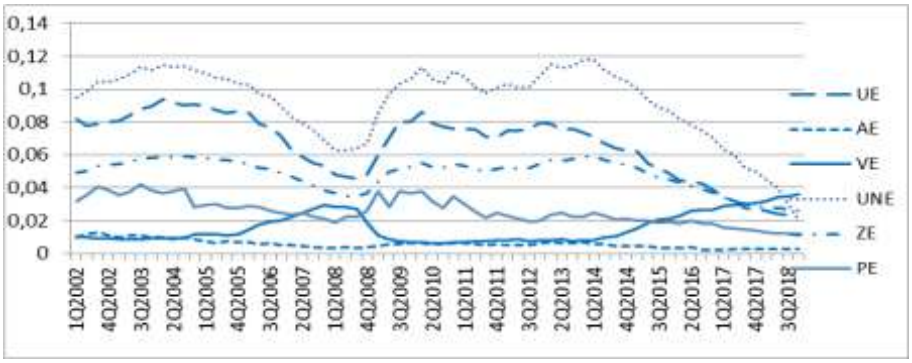


Figure 1 Variables

The testing of the bivariate and expanded labor market model was performed on seasonally adjusted quarterly time series for the period 2002–2018. Relationship (10) defines the ratios of the time series, which are used in the next text:

$$UNE_t = \frac{UN_t}{E_t}; \quad VE_t = \frac{V_t}{E_t}; \quad PE_t = \frac{P_t}{E_t}; \quad AE_t = \frac{A_t}{E_t}; \quad ZE_t = \frac{Z_t}{E_t}. \tag{10}$$

Given the fact that the original model of equation (3) is exponential, we will use transformed time series by means of the natural logarithm to examine the long-term relationship in the labour market in the given model. All variables were adjusted by logarithmic transformation. We define: LUNE = ln(UNE); LVE = ln(VE); LAE = ln(AE); LPE = ln(PE); LZE = ln(ZE).

The preparatory phase of estimating the VAR model is testing the stationarity of variables included in the model or their first differences. The interesting information about the inference in linear time series with some unit roots shows Sims *et al.* [17].

The test results for all variables are provided in Table 2. The Dickey-Fuller test (ADF) was used to test the stationarity. The second column provides information on the model type of testing the unit root (n = no trend and level constants /c = constant /c+t = level constant and trend), the third column contains the calculated T-statistics; the following column contains the corresponding level of statistical significance. The last column includes the result of testing: N = non-stationary (H0 not rejected), S = stationary (H0 rejected).

Variable	n/c/c+t	T-stat	Signif.	Result	Variable	n/c/c+t	T-stat	Signif.	Result
LAE	c+t	– 2.578	0.291	N	D(LAE)	n	– 1.611	0.049	S
LPE	c+t	– 2.998	0.141	N	D(LPE)	c	– 9.584	0.000	S
LUNE	c+t	– 1.661	0.756	N	D(LUNE)	c	– 3.791	0.005	S
LVE	c+t	– 2.598	0.282	N	D(LVE)	c	– 3.828	0.004	S
LZE	c+t	– 1.952	0.615	N	D(LZE)	n	– 3.177	0.002	S

Table 2 Testing the unit root of the variables in levels and their first differences

If the time series are non-stationary and cointegrated, the Vector Error Correction Model (VECM) can be used to examine the relationship of variables, as shown in [18]. VECM is the VAR model in the first differences complemented by the cointegration residue vector.

3 The long-term relationship in the labour market

3.1 Testing the long-term relationship in the labour market - 2 variables

On the basis of the obtained results, a cointegration analysis can be made using the Johansen test. This chapter will focus on the bivariate model and will examine the long-term relationship between *LUNE* and *LVE* for the Czech Republic. It confirms the existence of two long-term relations with a level constant and no trend in the cointegration equation. But there are two variables and it means that there is only one cointegration vector. Existence of one long-term bond with a level constant and no trend in the cointegration equation can be specified by a cointegration equation:

$$EQ2 = LUNE + 0.151LVE + 3.0028 \tag{11}$$

A cointegration vector expressing the relationship between *LUNE* and *LVE* is (1.000; 0.151; 3.0028). This means that a 1% increase in *VE* will cause a decrease in *UNE* by 0.151%. This conclusion is consistent with the premise, because a negative relationship is assumed between the variables.

$$D(LVE) = 0.476 \cdot [LUNE(-1) + 0.151 \cdot LVE(-1) + 3.0028] - 0.737 \cdot D\big(LUNE(-1)\big) + 0.222 \cdot D\big(LVE(-1)\big) + 1.441 \tag{12}$$

The coefficient of error correction factor EC1 (0.476) in the vacancy rate equation is statistically significant, which confirms that the variable of unemployment rate is causally linked to the job vacancy rate. An important coefficient of the correction factor proves the model's ability to explain short-term dynamics and convergence to long-term.

The Granger causality test indicates that changes in *LVE* precede changes in *LUNE* by 1, 2, 3 and 4 quarters. Thus changes in *LVE* can be used as predictor of short-run trend in *LUNE* movement. And changes in *LUNE* precede changes in *LVE* by 4 quarters, it means 1 year.

3.2 Testing the long-term relationship in the labour market - 3 variables

Testing results of the cointegration relationship of *LUNE*, *LVE*, *LPE* variables: that there is a long-term relationship in the labour market in the Czech Republic with inclusion of the *LPE* variable (the rate of registered unemployed applicants earning unemployment benefits). However, this model does not fulfil the assumption that the *LUNE* time series will develop in the long run in an inversely proportional relationship with the *LVE* time series.

Testing results of the cointegration relationship of *LUNE*, *LVE*, *LAE* variables: there is not a long-term relationship in the labour market in the Czech Republic with inclusion of the *LAE* variable (the rate of registered graduates among job applicants).

Testing results of the cointegration relationship of *LUNE, LVE*, *LZE* variables: that there is a long-term relationship in the labour market in the Czech Republic with inclusion of the *LZE* variable (the rate of registered unemployed women at labour offices).

It follows from cointegration equations that the *LUNE* time series develops inversely proportional to the time series *LVE* and directly proportional to the structural variable *LZE* in the long run. The coefficients of error correction factor EC1= –0.756, EC2 = –0.047 in the unemployment rate equation are statistically significant which confirms that the variable the job vacancy rate and the rate of registered unemployed women at labour offices are causally linked to the unemployment rate. The important coefficients of the correction factor prove the model's ability to explain short-term dynamics and convergence to long-term equilibrium. The values of the adjustment coefficient (–0.756 resp. –0.047) indicate that more than 75.6 % resp. 4.7 % of short-term deviations from the long-term equilibrium are corrected already in the following quarter. In other words, removing the complete imbalance would take approximately 4 months (1/0.756 = 1.3 quarter) for the job vacancy rate and more than 5 years (1/0.047 = 21.2 quarter) for the rate of registered unemployed women at labour offices. Regarding regression coefficients, it can be argued that the unemployment rate is negatively related to the job vacancy rate, with a quarterly delay. And the unemployment rate is positively related to the rate of registered unemployed women at labour offices and it is in line with the assumptions.

4 Conclusion

Job vacancy rate in relation to the unemployment rate reflects the labour market situation. A general rule is that the higher is the number of vacancies, the lower is the unemployment rate. Graphically, this rule is displayed by a Beveridge curve. This curve is used as one of the instruments for a labour market analysis in Eurostat.

This paper analysed the impact of the unemployment rate on the rate of job vacancies in the Czech Republic. Its purpose was to confirm or reject two hypotheses: H1: There is a long-term relationship between the unemployment rate and the rate of job vacancies in the Czech Republic, H2: There is a short-term relationship between the unemployment rate and the rate of job vacancies in the Czech Republic. Both hypotheses have been accepted. In the long-term period, LUNE (the unemployment rate) is the cause of a change in LVE (the rate of job vacancies). In the short-term period, it has been shown that changes in LUNE precede the change in LVE by 4 quarters.

Johansen's test confirms the existence of two long-term relations (with a level constant and no trend in the cointegration equation) between the unemployment rate and the rate of job vacancies. The value of the adjustment coefficient (0.476) in the equation (12) indicates that more than 47.6% of short-term deviations from the long-term are corrected already in the following quarter. In other words, removing the complete imbalance would take approximately half a year (1/0.476 = 2.1 quarter). Regarding regression coefficients, it can be argued that the job vacancy rate is negatively related to the unemployment rate, with a quarterly delay, i.e. a drop in the unemployment rate is followed by a rise in the job vacancy rate after one quarter. And this is in line with the assumption that a negative relationship is assumed between the variables.

In case of the trivariate model, existence of a long-term relationship was only demonstrated by inclusion of a variable in the rate of unemployed women, but in this case the residues are burdened with heteroscedasticity. A long-term relationship has been demonstrated in case of the trivariate model with inclusion of the variable rate of registered job applicants earning unemployment benefits, but the assumption that the *LUNE* time series has developed in the long run in proportion to the *LVE* time series has not been met. Existence of a long-term relationship on the labour market in the Czech Republic was not established in case of the trivariate model with inclusion of the variable of the rate of registered graduates among jobseekers.

It can be stated that there is a long-term relationship between the change in the unemployment rate and the change rate of job vacancies in the Czech Republic. Furthermore, the long-term relationship between the changes in the unemployment rate, the rate of job vacancies and the rate of unemployed women was confirmed. Results in this article do not match the conclusions in Hančlová *et al.* [13], where existence of a long-term relationship with inclusion of a structural variable was confirmed, i.e. the variants of the model (*LUNE*, *LVE*, *LPE*), (*LUNE*, *LVE*, *LAE*) and (*LUNE, LVE*, *LZE*) for the period 1995–2001. The reason for the discrepancy may be the time period. This article dealt with a data analysis for the period 2002–2018 and achieved empirical results are influenced by the fact that the Czech economy has undergone the period of currency crisis.

The benefit of this research is that the article confirmed cyclical unemployment, because reducing (increasing) the number of unemployed with increasing (declining) job vacancies is associated with cyclical changes. The cyclical unemployment was confirmed for model with inclusion of the variable the rate of registered unemployed women at labour office. Current movements of both variables in the same direction would point to structural unemployment, which was confirmed for model with inclusion of the variable the rate of registered unemployed applicants earning unemployment benefits in the Czech Republic in the period 2002-2018. The financial motivation to work is relatively low in the Czech Republic because it reflects the setting of taxes and social benefits. The results of the article show that the relationship between the minimum wage and unemployment benefits is an important factor.

Acknowledgements

This paper was supported by the project no. SGS/19/2019, Application of Customer Relationship Management Systems in Small and Medium-sized Enterprises accepted in 2019.

References

[1] Abe, Y. (2013). Regional variations in labor force behavior of women in Japan. *Japan and the World Economy 28*, 112–124.
[2] Al-Ubaydli, & O., List, J. A. (2016). Field experiments in markets. *National Bureau of Economic Research Working Paper Series*, No. 22113.
[3] Anthony, J.D.F. (1999). The relationship between unemployment and vacancies in Autralia. *Applied Economics 31*, 641 – 652.
[4] Bewley, R.A. (1979). The dynamic behaviour of unemployment and unfilled vacancies in Great Britain: 1958 – 1971, *Applied Economics* 11, 303 – 308.
[5] Blanchard, O.J.,& Diamon, P. (1989). The Beveridge Curve. *Brookings Papers on Economics Activity*,1-76.
[6] Bleakley, H., & Fuhrer, J.C. (1997). Shifts in the Beveridge Curve, Job Matching and Labor Market Dynamics. *New England Economic Review* 7.
[7] Čabla, A., & Malá, I. (2017). Modelling of unemployment duration in the Czech Republic. *Prague Economic Papers*, 26(4), 438–449.
[8] Doležalová, G. (2014). *Analýza volných pracovních míst evidovaných Úřadem práce ČR v kontextu se strukturou absolventů a zaměstnaných*. Retrieved from: http://www.nuv.cz/uploads/Vzdelavani_a_TP/Volna_mista_UP_2014_pro_www.pdf
[9] Filer, R. K. & Hamermesh, D. S. (1996). *The Economics of Work and Pay.* In Rees, A. E. (Eds.). 6th ed. New York: Harper Collins.
[10] Flek, V., & Mysíková, M. (2015). Uneployment dynamics in Central Europe: a labour flow approach. *Prague Economic Papers*, 24(1), 73–87.
[11] Gündüz, U., & Kaya, T. (2017). Regional employment generation potential of the Turkish labor market: an inter-sectoral perspective. *Technological and Economic Development of Economy*, 23(5), 726–741.
[12] Hančlová, J., & Šimek, M. (2001). Vliv vybraných charakteristik trhu práce na vývoj nezaměstnanosti v ČR. *Statistika* 2/2001, (38), 4 – 8.
[13] Hančlová, J., & Křivý, I.(2002). *Modelování a klasifikace regionálních trhů práce*. In Gottvald, J., Šimek, M., Tvrdík, J., Tvrdý, L. (Eds.). Ostrava: VŠB - Technická univerzita Ostrava, Ekonomická fakulta. Retrieved from: http://accendo.cz/wp-content/uploads/2002_MKL_RTP.pdf
[14] Hudcovský, M., & Lábaj, M. (2017). Employment growth and labour elasticity in V4 countries: Structural Decomposition analysis. In K. Morvay (Eds.). *Prague Economic Papers*, 26(4), 422–437.
[15] MPSV. (2017). *Statistiky nezaměstnanosti. Struktura uchazečů.* Ministry of Labor and Social Affairs. Retrieved from: http://portal.mpsv.cz/sz/stat/nz/qrt
[16] Pan, L. (2014). The impacts of education investment on skilled–unskilled wage inequality and economic development in developing countries, *Economic Modelling* 39, 174–181.
[17] Sims, C. A. & J. H. Stock. (1990). Inference in Linear Time Series Models with Some Unit Roots. In M.W. Watson (Eds.). *Econometrica*, 58(1), 113–144.
[18] Stoklasová, R. (2018). Default rate in the Czech Republic depending on selected macroeconomic indicators. *E&M Economics and Management*, 21(2), 69–82.
[19] Škare, M., & Caporale, G. M. (2014). Short- and long-run linkages between employment growth, inflation and output growth: evidence from a large panel, *Technological and Economic Development of Economy*, 20(3), 554-575.

Sensitivity Analysis of Priorities of Project Team Roles Using the ANP Model

Jan Rydval[1], Jan Bartoška[1], Tereza Jedlanová[1]

Abstract. The article describes the use of sensitivity analysis in the Analytical Network Process (ANP) for analysing how the priorities of the project team roles change when changing the priorities of individual project documents.
Model of Analytical Network Process is used for prioritization of the elements of the soft structure of projects. This soft structure is created using the semantic networks, which are based on the Work Breakdown Structure and the RACI matrix. Prioritization of the soft structure elements, of the project team roles in communication or in documents creation in the project management, is the key to the successful project completion. Sensitivity analysis of the project team roles seems to be an essential part of approaches to increase the success of a project. The results of the sensitivity analysis show the stability of the weights of the individual roles due to the change in weights of the individual project documents.

Keywords: Analytic Network Process, Project Team Roles, Project Management, Sensitivity Analysis, Supermatrix

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Working under specific conditions and under a specific timetable needs all team members to be well-informed. This is very important when working on a project. If a problem arises in one part of a project, it can mostly negatively impact other parts as well, and finally the project can fail. Communication is in this case the best way to avoid problems. According to the Project Management Institute (PMI), most project failures are due to communication issues. Proper communication plan helps to make sure that the right messages are sent, received, and understood by the right people including the people in a project team [12]. To set up appropriate and effective communication within the project team, it is necessary at first to define the project environment in a company. This can be done, for example, by creating a semantic model, as reported by Bartoška [2], and by El-Gohary, Osman, and El-Diraby [5], also according by Rydval, Bartoška and Brožová [15] semantic networks are suitable for displaying and expressing management structures and processes. The soft structure of semantic model is created using the semantic networks. These are based on the Work Breakdown Structure, the oriented hierarchical decomposition of works and parts of project management environment in the company, and the RACI matrix which is mostly used for clarifying and defining roles and responsibilities in the company. Brenan [3] explains how to use it to obtain information about the key people for communication and to split them into four groups: Responsible, Accountable, Consulted, and Informed.

After the project management environment is defined, i.e. it is known what organizational elements are important, what documents are most used, what are the limits and strategies of the project management environment. It is necessary to determine the weights (global priorities) for individual team roles. Then it will be possible to customize the communication plan to match the project environment and to ensure the success of the project. A semantic network can be converted to a network consisting of clusters and nodes for analytic network process to determine the cardinal quantitative information about the alternatives, with respect to the network structure of the decision problem. Saaty ([17], [18]) describes this issue quite in detail and he describes the dependency among decision elements. Prioritization of the soft structure elements especially project team roles in communication, is then the key to the successful project completion. Then it is necessary to test the stability of weights of elements, i.e. the stability of values of global priorities of network elements (especially team roles). This means whether their priorities change their value when changing the value of the criteria or other elements of the network structure of the project management environment. Nurdianaa, Wibowoa and Hatmokoa [11] have pointed out the need for sensitivity analysis when they conducted a sensitivity analysis to determine the risk for

[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

individual stakeholders using the AHP to define how the risk would change when criteria are changing. Similarly, Ziemba [21] recommends sensitivity analysis in his research, where he solves the problem of selecting the location and design of a wind farm using AHP and ANP. In his research, it has been found out that the results obtained with the use of the ANP is characterized by a higher quality than of the AHP. De Jesus, Gomes and Filardi [4] used ANP sensitivity analysis to determine the robustness of the results of the decision-making problem. Sensitivity analysis is thus suitable for assessing the change of preference of alternatives when the values of the other elements of the ANP network are changing.

The aim of the paper is to evaluate priority of the project team roles and test the stability of the global priorities of project team roles in the relation with the project document priorities. The sensitivity analysis of priorities (weights) is made by the ANP model describing the project managers' views on project management environment in a commercial unit by creating. The results of the sensitivity analysis show the stability of the weights of the individual roles due to the changes in weights of the individual project documents.

2 Materials and methods

The project management environment is a very complex system which cannot be structured hierarchically. It cannot be described by the AHP model ([16]), because it involves many interactions and dependences of higher-level elements in a hierarchy on lower level elements. Therefore, the ANP model represented by a network has to be used ([17], [18]). Its network in this case is based on the semantic network of described problem.

2.1 Analytic Network Process

The Analytic Network Process (ANP) is a generalization of the Analytic Hierarchy Process (AHP). The ANP model can reflect the increasing complexity of a network structure, where the network is created from different groups of elements and includes the dependences between the elements of the hierarchy. Each group of elements (the network cluster) consists of a homogeneous set of elements. Connections can exist between clusters as well as between the elements i.e. between the elements inside the cluster as well as between the elements from different clusters. The 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. Then the Unweighted and Weighted Supermatrices are created. The synthesized global priorities are received by the Limit Supermatrix calculation ([17], [18]). Raising the Weighted Supermatrix to powers generates the Limit Supermatrix. These powers converge directly to the Limit Supermatrix, or to a cycle of matrices and the Limit Supermatrix is the average of these matrices ([1]). From Limit Supermatrix the final global preferences of network elements are obtained.

Sensitivity analysis in the ANP and SuperDecisions software

Sensitivity analysis is recommended to check the stability of global preferences (elements ranking) obtained through ANP model ([6]). The basic steps of the sensitivity analysis of ANP are following:

- Select the cluster for sensitivity analysis of its elements (cluster of alternatives). The decision problem structure is shown as a network; therefore, any cluster can theoretically be determined as a cluster of alternatives. However, it is necessary to select a cluster that matches the purpose of the decision problem.
- Then, after changing the weights of the selected nods (elements of the non-alternatives clusters) in the Unweighted Supermatrix, the Limit Supermatrix is re-calculated and the new preferences of alternatives are received.
- Gradually increasing (from 0 to 1) the importance of a particular nod, the preferences of each alternative are re-calculated. This is way how to perform a sensitivity analysis of preferences of alternatives related the change of a specific nod. The obtained results of ANP are tested to stability (the robustness of the results when changing the rank of the specific nods).

All calculation of sensitivity analysis in the ANP can be provided via SuperDecisions software [20].

2.2 Semantic Model, WBS, RACI matrix

A semantic model consists of a semantic (associative) network that Mařík [10] defines as "natural graph representation". In the semantic network each node represents individual objects of described world and edges connecting these nodes and represents relationships between these objects ([19]). The term "semantic network" was for the first time used by Quillian [13] in his dissertation on the representation of English words and according

to Rydval, Bartoška [14] and Rydval, Bartoška and Brožová [15] are semantic networks suitable for displaying and expressing big information resources, management structures and processes or other areas.

The Project Management Body of Knowledge (PMBOK) defines the work breakdown structure as a "deliverable oriented hierarchical decomposition of the work to be executed by the project team." It is a key project deliverable that organizes the team's work into manageable sections. See [12] for information to WBS theory. A responsibility assignment matrix ([12]), also known as RACI (Responsible, Accountable, Consulted, and Informed) matrix is mostly used for clarifying and defining roles and responsibilities in cross-functional or departmental projects ([3]). See [12] for information to RACI matrix theory.

3 Results and Discussion

This case study shows how ANP sensitivity analysis can be used to determine the importance (weights) of the roles of a project team and how these changes become dependent on changes of the priorities of individual project documents. The case study follows the previous research ([2]) in the chosen bank organization (commercial unit), which took place from 2016 to 2018. Within the research, a basic semantic model of project management was created and described, the model includes a complete network of project roles, departments, project documentation, project restrictions, etc. Information about the project environment from project managers on different levels of the project office were obtained via a questionnaire survey.

3.1 Case study: ANP model of a commercial unit

The ANP structure of project management environment (Figure 1) in the commercial unit was created according information obtained from the previous research ([2]), i.e. from the semantic networks, which are based on the WBS and the RACI matrix.

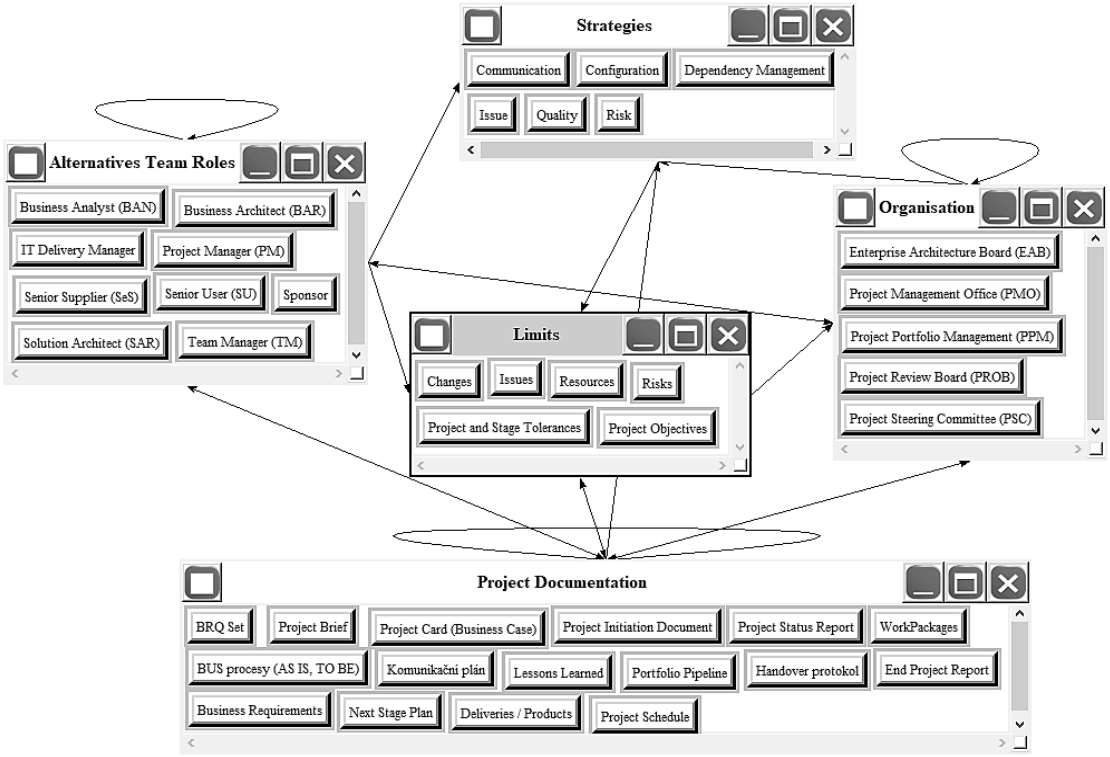


Figure 1 ANP model of project management environment in commercial unit (created using SuperDecisions)

The project management environment consists of five clusters: Strategies, Organisation, Project Documentation, Limits, and team roles (cluster of alternatives) as shown in Figure 1. The nine main project team roles are defined: Business Analyst (BAN), Business Architect (BAR), IT Delivery Manager (ITDM), Project Manager (PM), Senior Supplier (SeS), Senior User (SU), Solution Architect (SAR), Sponsor (S), Team Manager (TM). In all pairwise comparisons the model elements are supposed to be of equal importance, what means that all nodes are equal to each other. The global preferences of nodes are only dependent on the structure of the model. No personal subjective preferences of nodes are included. Prioritization of the project team roles is necessary to precise the project communication plan. The preferences are shown in three ways (Table 1).

Name	Ideals	Normals	Raw
Business Analyst (BAN)	0.000228	0.000185	0.000027
Business Architect (BAR)	0.001235	0.001003	0.000148
IT Delivery Manager	0.120378	0.097726	0.014386
Project Manager (PM)	1.000000	0.811828	0.119509
Senior Supplier (SeS)	0.034156	0.027729	0.004082
Senior User (SU)	0.034156	0.027729	0.004082
Solution Architect (SAR)	0.000051	0.000041	0.000006
Sponsor	0.007411	0.006017	0.000886
Team Manager (TM)	0.034172	0.027742	0.004084

Table 1 Quantification of sub-roles in the project team (authors calculation using SuperDecisions)

The Raw column contains the values of preferences from the Limit Supermatrix, the Normals column shows the value of preferences normalized for Team Roles cluster, and the Ideals column shows the preferences obtained by dividing the values by the largest value in the column. In the project management environment in the commercial unit, the Project Manager (PM) is evaluated as the most important role. However, since the importance of single documents may change within the project environment, especially at different stages of the project, it is useful to know how individual role preferences will change when changing the importance of individual documents.

3.2 Sensitivity Analysis of priorities of Project Team Roles

Creation of the precise project communication plan includes testing the value of individual role preference. The individual value of roles were tested using the ANP sensitivity analysis. This was performed by changing the preference of individual project documents, especially Project Brief, Stage Plan, Work Packages, and Project Status Report.

The results of the sensitivity analysis of project role preferences is shown in the following graphs. The x-axis of each graph represents the value of the document's preference $[0, 1]$, and the y-axis represents the changing importance of each project role $[0, 1]$. The first two charts (on the top of the Figure 2) show the sensitivity analysis results for the change in importance of the documents Project Brief and Stage Plan. It is clear that when changing the preference of these two documents, the role of the PM remains much more important than the other roles of the project team. Even if the importance of these two documents increases significantly, the role of PM is even more important. On the other hand, when the preference of the Work Packages is rising above 0.5, the importance of the PM is declining in favor of the Team Manager (on the bottom left of the Figure 2). If the preference of Status Report rises above 0.5, the importance of the PM decreases in favor of the Senior User and Sponsor, but it is still the most important role. It is thus evident that the role of PM does not remain the most important at all times, but in certain cases the importance of other roles can also increase. It is therefore necessary to adapt the plan of project communication and to take these facts into account.

It should be pointed out, that results of team roles global preferences as well as results obtained from the sensitivity analysis in this paper refer only to the specific management environment obtained by ANP methods. As Ziemba [21] states in his work, it should be clear, that without further research, the values of results cannot be generalized; however, it is evident that the values of role importance could be unstable. Therefore, the main benefit of this paper is a demonstration of the practical application of how to determine the stability or instability of elements of the project elements environment by using ANP sensitivity analysis due to changes in other network elements. The knowledge of the stability of the preferences of the project team roles depending on changes of project documentation is very important and it has to be involved into the creation of project communication plan or to the others stakeholder's analysis.

In this research, the sensitivity analysis shows that the global preferences of particular project team roles are not very sensitive to the project documentation changes. The importance of roles is relatively stable, except for Project Manager and Team manager. Therefore, the decision-makers can effectively estimate the suitable communication plan. This sensitivity analysis was conducted only due to the change in weights of the individual project documents. For more precise communication plan the sensitivity analysis due to the change in weights of the individual nods of other cluster can be involved. Also, in the case, when the decision makers can decide the importance of ANP structure elements, they can be sometimes inconsistent while filling in the pairwise comparison matrix.

Figure 2 ANP sensitivity analysis of project team roles (authors calculation using SuperDecisions)

The consistency of the matrix then reaches over a feasible consistency limit and the information value of the data can be ruined ([8]). The issue of evaluation of project team role importance shows such complexity, that the application of the ANP is appropriate, Ziemba [21] also pointed out the higher quality of ANP results. For the decision problems containing the dependence between the elements of the same level of hierarchy is ANP more suitable approach than AHP. Furthermore taking into consideration that real dependencies between criteria in the decision model make the model precisely reflect the real decision problem and it allows to obtain more precise results ([7], [9]).

4 Conclusion

Proper communication plan is an important component of project success, and it ensures that projects does not fail du to communication issues. However, for a proper communication, the importance of project team roles has to be evaluated, and the stability of such evaluation has to be estimated. This article describes the use of the ANP sensitivity analysis for analyzing how the importance of the project team role changes when changing the priorities of individual project documents. The practical application of the ANP sensitivity analysis to determine the stability of importance of project team roles is the main benefit of this paper. In this paper the ANP model of project management environment in a commercial unit was constructed. After the project team roles quantification the sensitivity analysis of importance (weights) of project team roles is accomplished. The results of the sensitivity analysis show the stability of the weights of the individual roles due to the change in weights of the individual project documents. In this case study the following roles: Business Analyst, Business Architect, IT Delivery Manager, Senior User, and Solution Architect are not very sensitive to the project documentation changes, and the roles: Project Manager, Senior Supplier, Sponsor, Team Manager are slightly sensitive to the project documentation changes.

Acknowledgements

Acknowledgements This research is supported by the grant No. 2019A0015 “Ověření a rozvoj sémantického modelu řízení projektů” of the Internal Grant Agency of the University of Life Sciences Prague

References

[1] Adams, B. (2011). *SuperDecisions Limit Matrix Calculations*. Decision Lens Inc.
[2] Bartoška, J. (2016). Semantic Model of Organizational Project Structure. *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Technical University of Liberec, 19-24.
[3] Brennan, K. (2009). *A Guide to the Business Analysis Body of Knowledge (BABOK Guide)*. International Institute of Business Analysis.
[4] de Jesus, V. M., Gomes, L. F. A. M. & Filardi, F. (2019). The Selection of Oil & Gas Projects Contract Strategies with the ANP. *Independent Journal of Management & Production*, 10 (2), 355-379.
[5] El-Gohary, N. M., Osman, H. & El-Diraby, T. E. (2016). Stakeholder management for public private partnerships. *International Journal of Project Management*, 24, 595-604.
[6] Farman, H., Javed, H., Jan, B., Ahmad, J., Ali, S., Khalil F., N., Khan, M. (2017). Analytical network process based optimum cluster head selection in wireless sensor network. *Future Generation Computer Systems ONE*. 12 (7).
[7] Golcuk, I. & Baykasoglu, A. (2016). An analysis of DEMATEL approaches for criteria interaction handling within ANP. *Expert Syst. Appl*, 46, 346–366.
[8] Hlavatý, R. (2014). Saaty's matrix revisited: Securing the consistency of pairwise comparisons. *Proceedings of the 32th International Conference Mathematical Methods in Economics*. Palacký University, Olomouc, 287-292.
[9] Mandic, K., Bobar, V. & Delibasi´c, B. (2015). Modeling Interactions Among Criteria in MCDM Methods: A Review. Lect. *Notes Bus. Inf. Process*, 216, 98–109.
[10] Mařík, V. (1993). *Umělá inteligence: Díl 1*. Praha: Academia.
[11] Nurdianaa, A., Wibowoa, M. A. & Hatmokoa, J. U. D. (2015). Sensitivity analysis of risk from stakeholders' perception Case study: Semarang-Solo highway project section I (Tembalang-Gedawang), *Procedia Engineering 125*. 12–17.
[12] Project Management Institute (2013). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. Project Management Institute Inc., 14 Campus Boulevard.
[13] Quillian, R. (1968). A recognition procedure for transformational grammars. Doctoral dissertation. Cambridge (MA): Massachusetts Institute of Technology.
[14] Rydval, J. & Bartoska, J. (2013). Quantification of Framing Effect in the Meat Distribution by ANP. *Proceedings of the 31st International Conference Mathematical Methods in Economics*. College of Polytechnics Jihlava, Jihlava,779-784.
[15] Rydval, J., Bartoška, J., & Brožová, H. (2014). Semantic Network in Information Processing for the Pork Market. *AGRIS on-line Papers in Economics and Informatics* 6, 59-67.
[16] Saaty, T. L. (2000). *Fundamentals of the Analytic Hierarchy Process*. Pittsburgh: RWS Publications.
[17] Saaty, T. L. (2001). *Decision Making with Dependence and Feedback: The Analytic Network Process*, The Analytic Hierarchy Process Series. Pittsburgh: IX, RWS Publications.
[18] Saaty, T. L. (2003). *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback,* Creative Decisions Foundation.
[19] Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Pacific Grove (CA): Brooks/Cole Publishing Co.
[20] *SuperDecisions Software for Decision-Making*. (2018). URL <http://www.superdecisions.com/>.
[21] Ziemba, P. (2019). Inter-Criteria Dependencies-Based Decision Support in the Sustainable wind Energy Management. *Energies 2019*, 12, 749.

The identification of the Key players in the supply chain network

Jaroslava Pražáková[1], Martin Pech[2]

Abstract. Regarding to the future challenges, another aspect of supply chains needs to be mentioned. Network design decisions have a significant impact on performance and flexibility within the supply chain networks. In this paper, we used various methods suitable for external analysis of automotive industry network that can identify the key players of the network. Exponential Non-Discriminative Snowball Sampling was used as a method for network construction; existence of each 70 nodes and 110 links (based on information flow) was verified using the real suppliers or customers relationship. Then, the analysed network is specified by the basic whole network measures. The results of whole network measures clue method selection used in the next step of the network analysis. The core and periphery was identified using the Core Periphery Model. Within the core part of the network, key players were divided into hubs and authorities through other network characteristics of closeness, information centrality, etc. The node fragmentation level determines the importance of each node for network structure and financial stability.

Keywords: supply chain network, network analysis, core periphery model, fragmentation, network stability

JEL Classification: L14, L60
AMS Classification: 90B10, 90C35

1 Introduction

Given the rapidly changing environment and the challenges that supply chains will face in the foreseeable future, other aspects affecting supply chains are emerging. The relatively recent incorporation of the term “network” into supply chain management research represents a pressing need to view supply chains as a network for firms to gain improved performance, operational efficiencies, and ultimately sustainable competitiveness [15].

Network analysis draws on theories from the social, organizational, and complexity sciences and leverages graph theoretic methods to model, analyse, and visualize the structure, dynamics, and strategies that shape supply chains [2]. A supply chain can be modelled as a network by a set of “nodes” that represent autonomous business units as firms that are able to exercise sovereign choices, and a set of “connections” (ties) that link these firms together for the purposes of creating products or services [9]. Network structure can therefore be associated with the number, identity, and characteristics of nodes; the location, content, or strength of ties; and the pattern of interconnections or ties among nodes [1]. The relationships between and among two or more firms, in aggregate, define the overarching structure of the network in which a firm exists [5]. The control of information and information as a source of power are seen significant, especially on the relationship and network levels [6]. The main barriers to supply chain information integration are lack of supply chain information leadership by the larger firms, lack of awareness of the benefits of eBusiness, particularly in the smaller upstream businesses [8].

According to the basic theory, the supply network is usually centrally managed by one large enterprise, called a "focal enterprise", “head player”, “key player”, “key enterprise”, a network integrator who has a leading role in supply chain management. Then, the key player in a supply chain is the member, who overlooks and leads others, creates an environment in which the partnership with suppliers can grow accordingly [11]. The presence

[1] University of South Bohemia in České Budějovice /Faculty of Economics, Department of Accounting and Finances, Studentská 13, České Budějovice, Czech Republic, smoloj@ef.jcu.cz
[2] University of South Bohemia in České Budějovice/Faculty of Economics, Department of Management, Studentská 13, České Budějovice, Czech Republic, mpechac@ef.jcu.cz

of key players such as hub firms in supply chain systems is significant. These firms, are said to control performance and provide system-wide coordination of the supply chain [10]. The presence of a power law connectivity distribution and hub firms is empirically observed in material flows in various automotive supply chains [13]. The network position is a structural determinant of power. The power depends on the network (roles, positions, number of actors, level of domination), relationship (technology, information, negotiation skills, volume of sales etc.) and organization characteristics (resources, brands, enterprise size).

Supply chain networks are highly vulnerable to disruptions. A disruption in a supply chain is an unplanned and unanticipated event that disrupts the normal structure of flows in supply chain structure [7]. Underestimating disruptive risks, for instance, by completely ignoring them, is a dangerous bet. Strategies designed to deal with disruption risk (such as having multiple suppliers) compensate for the upfront cost to some extent by providing some benefits even in the context of recurrent risk [12]. The network's robustness against these disruptions can be analyzed. Decision support systems allow users to evaluate performance of a supply chain network before and after given various types of disruptions [18]. Although scholarly understanding of the factors that influence the formation of relationships between entities exists, understanding the origins and evolution of alternative types of network structures remains a research issue demanding attention.

2 Material and methods

Paper is focuses on identification of the key network players and their roles based on the network analysis. We try to explore if key managing company is identical to network centre with respect to connectivity and centrality features. The idea is verified base on the real case study of automotive supply network. Building on the existing research gaps, this study attempts to answer two research problems:

1) Which nodes are key players in the supply chain network and what roles they play?
2) Is there a difference between enterprises in core and periphery of a network?

The network characteristics are used to show differences between core and periphery of the network. The obtained results were subjected to statistical analysis by t-test. The results are interpreted at alpha significance level 0.05, resp. with 95% confidence. For clarity, the text contains only significant results, including achieved level of significance (p-value). Evaluation of individual statistical tests was performed by Statistica software.

2.1 The data source and collection

The paper analyses the automotive industry network. Exponential Non-Discriminative Snowball Sampling was used as a method for network construction; existence of each 70 nodes and 110 links was verified using the real suppliers or customers relationship and oriented based on control information flow. At the very beginning of the network construction process, the questionnaire research performed in years 2007 – 2018 was used [16]. In the next step of the network construction, only the Exponential Non-Discriminative Snowball Sampling was used. The data from the company website as well as stated references were used. All companies in network are middle or big companies oriented on automotive industry. The original network was expanded by more than 50 automotive companies. Dataset is converted to a standard undirected unweighted representation of the network, a binary matrix consisted of 1s and 0s.

2.2 Methods and network measures

The selection of network variables and methods is based on several criteria. Because our study is network-based, it is important to identify constructs which define a network completely. Three groups of characteristics are used:

a) Basic network characteristics

- **Network size** - The number of nodes in a network.
- **Network density.** It is a ratio of total number of links and the total number of possible links.
- **Average degree**. The distribution of how connected individual nodes are describes network structure. Since the data are asymmetric (with directed links), we can distinguish between ties being sent (out-degree) and ties being received (in-degree).
- **Connectedness**. Differences among nodes can be used to understand how information moves in the network, which actors are likely to be influential each other, and a number of other important properties.
- **Fragmentation**. Fragmentation is the proportion of pairs of nodes that cannot reach each other. Distance weighted fragmentation is one minus the average reciprocal distance between all pairs of nodes.

b) Network centrality measures

- **Information centrality** (C_I) of a node *i* is the harmonic mean of all the information measures between *i* and all other nodes in the network. The information measure between two vertices *i* and *j* is the inverse of the variance of the weighted function [17].
- **Degree centrality** (C_D) is defined as the number of links adjacent to a given node (i.e., the number of links that a node has).
- **Eigenvector centrality** (C_E) defines a node's centrality as a weighted average of the centrality of its neighbors.
- **Closeness centrality** (C_C). Closeness centrality is usually interpreted either as a measure of access efficiency or of independence from potential control by intermediaries.
- **Betweenness centrality** (C_B). Betweenness centrality measures the number of times a node lies on the shortest path between other nodes. It is based on the idea that broker position between others provide the opportunity to intercept or influence their communication. Betweenness is generally employed with the understanding that it captures the potential for control of communication between actors [4].
- **Fragmentation centrality** (C_F) of a node is the difference in the total score including the node and the score with the node removed. Two measures of fragmentation centrality are used: change in network fragmentation when node is removed (C_{Fd}) and change in distance-weighted network fragmentation when node is removed (C_{Fdw}).

c) Network models

- **Core/periphery** discrete model uses a genetic algorithm to fit a core/periphery model to data and identify which actors belong to the core and which belong to the periphery. The fit is simply the correlation between the data matrix and an ideal structure matrix in which 1 means the core block interactions and a 0 represents the peripheral block interactions. The algorithm seeks the maxima of the cost function [3]. The model was computed via UCINET software.
- **Hubs and authorities** scores of a analysed network are based on natural generalization of eigenvector centrality. A high hub actor points to many good authorities and a high authority actor receives from many good hubs [14].
- **Key Players model.** Model identify an optimal set of nodes in a network for one of two basic purposes: (a) crippling the network by removing key nodes (fragmentation experiment), and (b) selecting which nodes to keep influence under surveillance after intervention. The two purposes are different and require different procedures. The model can be computed via KeyPlayer software.

3 Results

3.1 Network characteristics

The structure of any network can be conceptualized in terms of three primitives—the nodes that comprise the network, the ties that connect the nodes, and the patterns or structure that result from these connections. Network structure can therefore be associated with the number, identity, and characteristics of nodes; the location, content, or strength of ties; and the pattern of interconnections or ties among nodes [1]. From the original 18 network companies, the network has grown to its present size over almost 10 years. All the time we have been monitoring its development. Today network size, represented by number of nodes and links, incorporates 70 companies (nodes in Figure 1 as C and number), where we were able to reliably identify the 110 connections (links). All links are constructed base on control information flow between the nodes; the arrows show the direction of the transmitted information. The network includes 8 global automakers (C17; C18; C19; C20;C25; C38; C50; C57) and more than 30 engineering companies that supply parts for the final assembly of passenger cars (for example C8; C10; C9; C24; C23; C36). Other networked businesses include suppliers of windshields, lighting, electronic components, seat and so on.

The analyzed network (see Figure 1) is characterized by low network density (0.023), asymmetric on the level 0,041 and compactness (0,059). Average degree of the network reaches the value 1.571. Total connectedness of the network is 0.139. Other important measure used for network structure description is fragmentation. The network fragmentation reaches the value 0.861.

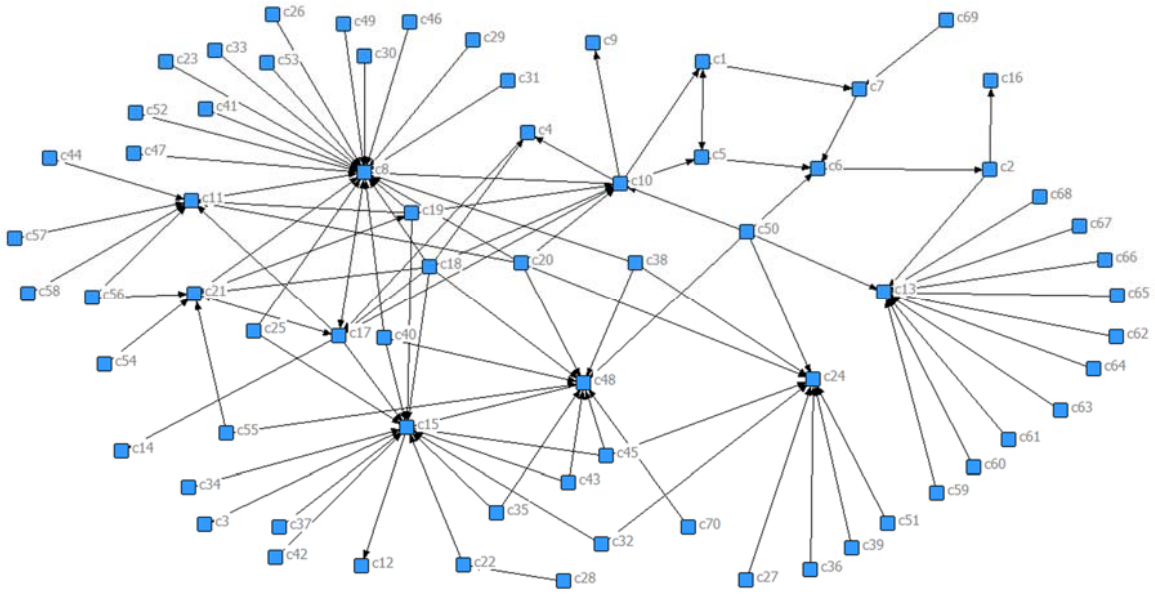


Figure 1 Supply chain network structure

3.2 Key players and their roles

Based on the results of types of centrality measures, we identified key companies of the supply network. Their roles are deeply analyzed.

The closeness centrality explains why one company is more powerful than others in network when this company is closer to more companies than any other. It can be used to find the best placed companies to influence the entire network most quickly. Their role can be called a navigator that explores, accesses, and collects information with greater autonomy in the supply network. We can distinguish two types of closeness:

- Out-closeness: C50 ($C_C = 0.144$), C18 ($C_C = 0.143$), C20 ($C_C = 0.142$), C10 ($C_C = 0.139$), C17 ($C_C = 0.138$)
- In-closeness: C15 ($C_C = 0.207$), C3 ($C_C = 0.188$), C12 ($C_C = 0.188$), C8 ($C_C = 0.177$), C11 ($C_C = 0.170$)

The degree centrality assigns an importance score based purely on the number of links held by each node. The key companies then can be called as coordinators. The role of coordinator can be explained as the ability to reconcile differences of network companies and align their own goal on the greater supply network goals. Two types of degree centrality can be measured:

- Out-Degree: C18 ($C_D = 7$), C17 ($C_D = 6$), C10 ($C_D = 6$), C19 ($C_D = 5$), C20 ($C_D = 5$)
- In-Degree: C8 ($C_D = 22$), C15 ($C_D = 14$), C13 ($C_D = 12$), C48 ($C_D = 10$), C24 ($C_D = 9$)

Betweenness centrality can be useful for analysing communication dynamics. A high betweenness value could indicate that someone holds authority over or controls collaboration between disparate clusters in a network, or indicate they are on the periphery of both clusters. Key companies in broker role (which means mediation of dealings between network members) are: C17 ($C_B = 373$), C8 ($C_B = 334.167$), C10 ($C_B = 272.667$), C5 ($C_B = 122$), C6 ($C_B = 104$).

Information centrality provides a more complex weighting of the distances between nodes, and it summarizes the closeness centrality of each node by the harmonic mean of its distances to the others. Key companies here are: C8 ($C_I = 0.916$), C15 ($C_I = 0.892$), C10 ($C_I = 0.889$), C48 ($C_I = 0.887$), C18 ($C_I = 0.857$).

Eigenvector centrality measures a node's influence based on the number of links and taking into account how well connected they are, and how many links their connections have, and through the network. Key companies were: C8 ($C_E = 0.454$), C17 ($C_E = 0.298$), C18 ($C_E = 0.296$), C15 ($C_E = 0.294$), C10 ($C_E = 0.286$).

The results show in particular the prevailing dominance of key companies C8 and C17. However, this dominance is no longer significant for other companies, as different types of centralities produce different results. In addition, a Core/periphery model would be used, that help to explore the network core.

3.3 Network models

We used categorical Core / periphery model, which identifies which nodes belong to the core and which belong to the periphery. The nature of the model is based on the idea that core nodes are adjacent to other core nodes, core nodes are adjacent to some peripheral nodes, and peripheral nodes do not associate with other peripheral nodes. The main measure of fit is based on the correlations with maximum 100 iterations and 5 random starts. The final core/periphery fit (correlation) is 0.403, which indicates average model fit. The results of the Core Class Memberships in blocked adjacency matrix are following: C8, C10, C11, C13, C15, C17, C18, C19, C20, C21, C24, C48 and C50. It means that about 18.57% companies belong to the core of the network.

The network characteristics of companies grouped in core and the periphery were then subjected to statistical analyzation by t-test in software Statistica. Results (table 1) show that the differences in all selected centrality measures are significant. The results of the means show that the values of the core companies are higher than values of the companies in the periphery.

	Mean		**Test of variance**		**Test of Means***	
Network characteristics	**Core**	**Periphery**	**F**	**p-value**	**t**	**p-value**
Eigenvector centrality	0.226	0.046	9.122	**0.000**	-5.838	**0.001**
Information centrality	0.833	0.501	2.160	0.144	-10.595	**0.000**
Betweenness centrality	95.001	5.930	35.153	**0.000**	-2.340	**0.037**
Out-closeness centrality	0.130	0.121	2.148	0.055	-3.220	**0.002**
In-closeness centrality	0.150	0.120	2.296	**0.037**	-3.398	**0.004**
Out-degree centrality	3.308	1.175	14.006	**0.000**	-3.097	**0.009**
In-degree centrality	7.077	0.316	89.5400	**0.000**	-3.753	**0.003**

Table 1 Differences in network characteristics between core and periphery (Statistica software)

*p-value results are two-tailed and selection of tests depend on the significance of the variance (t-test for equality of variances and non-parametric version of test for unequality of variances)

Another way to find key network players offer **software KeyPlayer** which used model based on the group centrality. The basic idea of the algorithm is to select a set of nodes as seeds and then swap the selected nodes with unselected ones if the swap increases the group centrality. The results of algorithm reach fragmentation index 0.891 and heterogeneity 0.879. The key companies according to this model are: C8, C10, C15, C24 and C48. However, the model is based on a predetermined number of key players that do not split into the core and periphery.

Hubs and authorities are “supporters” and “centres” in a bipartite core of a network. In general, there are more companies with hubs or authorities roles in the core than the periphery. This is confirmed by the results of the hubs and authority scores in Table 2. A good hub is one that links to many good authorities. The main hubs in core are companies C17, C19, C20. Relatively high hub scores have also companies C40 (0.667), C38 (0.559) and C25 (0.513) from the periphery. The most important authority is enterprise C8 with the highest authority score. It means that this company is one that links to by many good hub companies. No enterprise out of the periphery scores higher than companies from the core.

	C8	**C10**	**C11**	**C13**	**C15**	**C17**	**C18**	**C19**	**C20**	**C21**	**C24**	**C48**	**C50**
Hub Score	0.163	0.475	0.318	0.000	0.000	0.859	1.000	0.832	0.777	0.396	0.000	0.195	0.394
Auth. Score	1.000	0.378	0.305	0.059	0.613	0.206	0.000	0.206	0.000	0.318	0.274	0.484	0.000

Table 2 Hubs and Authority Score in network core (software R)

Fragmentation measures the extent to which a network is fragmented after a node is removed from the network. For the representation of this network characteristic, the difference between the original and the new fragmentation value (after removing the node) is most suitable. These measures and variables can be weighted according to the network density. The highest disruption to the network structure would arise after the removal of C17 (C_{Fd} = 0.080; C_{Fdw} = 0.023), C8 (C_{Fd} = 0.076; C_{Fdw} = 0.024), C10 (C_{Fd} = 0.064; C_{Fdw} = 0.017) and C15 (C_{Fd} = 0.023; C_{Fdw} = 0.0130). Results of fragmentation measures confirm the network dominance of C8, C10 and C17. In addition, the results also correspond to the functional network connection. C17 is one of the best-performing automaker with more than good financial power. Hence the network functionality would be severely damage should this company exit the network.

4 Conclusion

The paper deals with the identification of the key players in the network and their roles. The results show the dominant position of companies C8 and C17. The automaker C17 functions as broker, hub as well as coordinator. Furthermore, significant differences between core and periphery of the network were verified. Fragmentation indicators have also identified potential changes in the network structure, if key players are leaving.

Network dynamics and structure understanding is also important because of the potential role of network participants in creating network structures that benefit them. The key players in the network influence the long-term existence of the network and almost all network flows. The way how key players influence the network corresponds to their main role in the network and through it they are able to affect every day performance of the network.

References

[1] Ahuja, G., Soda, G. & Zaheer, A. (2012). Introduction to the Special Issue: The Genesis and Dynamics of Organizational Networks. *Organization Science,* 23 (2), 434-448.

[2] Bellamy, M. A. & Basole, R. C. (2013). Network analysis of supply chain systems: A systematic review and future research. *Systems Engineering,* 16 (2), 235-249.

[3] Borgatti, S. P. & Everett, M. G. (1999). Models of core/periphery structures. *Social Networks,* 21 (4), 375-395.

[4] Brandes, U., Borgatti, S. P. & Freeman, L. C. (2016). Maintaining the duality of closeness and betweenness centrality. *Social Networks,* 44 153-159.

[5] Carnovale, S., Rogers, D. S. & Yeniyurt, S. (2019). Broadening the perspective of supply chain finance: The performance impacts of network power and cohesion. *Journal of Purchasing and Supply Management,* 25 (2), 134-145.

[6] Cox, A. (2001). Understanding Buyer and Supplier Power: A Framework for Procurement and Supply Competence. *Journal of Supply Chain Management,* 37 (1), 8-15.

[7] Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J. & Handfield, R. B. (2007). The Severity of Supply Chain Disruptions: Design Characteristics and Mitigation Capabilities. *Decision Sciences,* 38 (1), 131-156.

[8] Harland, C. M., Caldwell, N. D., Powell, P. & Zheng, J. (2007). Barriers to supply chain information integration: SMEs adrift of eLands. *Journal of Operations Management,* 25 (6), 1234-1254.

[9] Hearnshaw, E. J. S. & Wilson, M. M. J. (2013). A complex network approach to supply chain network theory. *International Journal of Operations & Production Management,* 33 (3-4), 442-469.

[10] Human, S. E. & Provan, K. G. (2000). Legitimacy building in the evolution of small-firm multilateral networks: A comparative study of success and demise. *Administrative Science Quarterly,* 45 (2), 327-365.

[11] Chen, H. Y. S., Lin, C. W. R. & Yih, Y. (2007). Production-distribution network design of a global supply chain alliance from the key player's perspective. *International Journal of Production Research,* 45 (2), 245-265.

[12] Chopra, S. & Sodhi, M. S. (2014). Reducing the Risk of Supply Chain Disruptions. *Mit Sloan Management Review,* 55 (3), 73-+.

[13] Kim, Y., Choi, T. Y., Yan, T. T. & Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management,* 29 (3), 194-211.

[14] Kleinberg, M. J. (1999). Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM,* 46 (5), 604-632.

[15] Kotabe, M., Martin, X. & Domoto, H. (2003). Gaining from vertical partnerships: Knowledge transfer, relationship duration, and supplier performance improvement in the US and Japanese automotive industries. *Strategic Management Journal,* 24 (4), 293-316.

[16] Pech, M. & Smolova, J. (2010) Using of fuzzy entropy as a supportive method for managing the real supply chain: a case study. *Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010* (pp. 50-510). University of South Bohemia, Faculty of Economics, Ceske Budejovice.

[17] Stephenson, K. & Zelen, M. (1989). Rethinking Centrality - Methods and Examples. *Social Networks,* 11 (1), 1-37.

[18] Zhao, K., Scheibe, K., Blackhurst, J. & Kumar, A. (2019). Supply Chain Network Robustness Against Disruptions: Topological Analysis, Measurement, and Optimization. *Ieee Transactions on Engineering Management,* 66 (1), 127-139.

Least Squares Method With Equality Constraints and Polynomial Approximation of Lorenz Curve

Pavel Pražák[1], Kateřina Frončková [2]

Abstract. The least squares method is frequently and successfully used in the solution of various different econometric problems. No restrictions on data sets are usually considered. The only widely known problem with a constraint is the one where a linear model heading through the origin is solved. Nevertheless, we can encounter approximation problems having more data restrictions. For instance, the Lorenz curve which is a curve heading through two given points. In this case, it is useful to apply a least squares method subject to constraints. In this paper, two possible solutions of problems with natural data restrictions are examined. First, it is showed that the constrained problem with two boundary values can be transformed into the classical least squares problem and a special form of the normal equation is derived. A more general problem is then introduced and the Lagrange multiplier method is used to develop a different form of the normal equation. Finally, a polynomial approximation of the Lorenz curve applied to the Czech Republic income data is introduced.

Keywords: approximation, Lorenz curve, least squares method, normal equation, optimization.

JEL Classification: C44, C61
AMS Classification: 49J15, 49N90

1 Introduction

Fitting a functional model to experimental data and observation is one of the most fundamental problems of mathematical modelling. Moreover, data are subject to source of errors. The best solution can be determined in many different ways, [4], [3]. The least squares method is frequently used in case it is necessary to estimate parameters of a mathematical model from measured data, [4], [6]. Linear models and least squares methods that are not subject to any constraints are typically considered in different econometric problems [11]. These problems usually have a nonzero intercept but there are also problems where it is useful to consider zero intercept. Such problems are also known as regression through the origin, see [11]. Moreover, there are models that have a functional form with more general natural constraints. In this paper, we want to explore the least squares method for a polynomial approximation that considers linear constraints. The Lorenz curve that was first developed by Max O. Lorenz in 1905 can serve as a graphical representation of income distribution and also as a possible example of such a problem , cf. [9]. The least squares approximation of the Lorenz curve is examined in this paper. Such an approximation has to consider the least squares problem subject to constraints - points (0,0) and (1,1). It is shown that the problem can be transformed into a classical problem and a special form of the normal equation is derived in the paper. The vector of residuals of the least squares problem is used for a finding of the most suitable polynomial for approximation and its degree. The polynomial approximation of the Lorenz curve for the Czech Republic income data 2017 is used to demonstrate application of the presented least squares method with constrains.

2 Problem Description

One of the most frequently used problems in empirical research is to fit a given functional model to given observation data. The problem will be shortly described here, for details see [7]. Given a set of $m, m \in \mathbb{N}$, data points

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}, \tag{1}$$

consider they are to be fitted to a functional model

$$y = f(\beta, x), \tag{2}$$

[1] University of Hradec Králové, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, Czech Republic, pavel.prazak@uhk.cz
[2] University of Hradec Králové, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, katerina.fronckova@uhk.cz

where $\beta, \beta \in \mathbb{R}^{n+1}$, is a $(n+1)-$vector parameter to be found and $f : \mathbb{R}^{n+1} \times \mathbb{R} \to \mathbb{R}$. There are two weakness points connected with this assignment. First one lies in errors in the input data D (measurement errors) and second one in selection made in the particular functional model f (errors in model). We do not analyse these two types of problems in detail in the paper, for more details see [6], [4]. The simple case occurs if the functional model is linear in β and has the form of polynomial of the given degree n, $n \in \mathbb{N}$,

$$y = f(\beta, x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n. \tag{3}$$

Reasons why the class of polynomials of the given degree is selected are that polynomials and their derivatives are continuous functions and function values of polynomials can be easily found by the basic operations addition and multiplication. We consider the case when the number of observations (1) is greater than the number of parameters. Given $m > n$ the resulting equations with data set (1) for unknown vector parameter β are

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_n x_i^n, i \in \{1, 2, \ldots, m\}. \tag{4}$$

If the matrix $X \in \mathbb{R}^{m \times (n+1)}$ is defined as $X_{ij} = x_i^{j-1}$, where $i \in \{1, 2, \ldots, m\}$, $j \in \{1, 2, \ldots, n+1\}$ and the components of $y \in \mathbb{R}^m$ are defined as y_i, then the system (4) can be written in matrix form

$$X\beta = y. \tag{5}$$

This system of linear equations is overdetermined and its exact solution does not exist. The "solution" to linear system (4) can be defined in many possible ways. One of the usual way to find the best approximation is the least squares approximation, [7]. This is equivalent to finding minimum of the size of the residual vector $r(\beta) = y - X\beta$. Therefore, the optimal solution $\widehat{\beta}$ which minimizes the sum of squared residuals components is the solution to the optimization problem

$$\min \left\{ ||y - X\beta||^2 \mid \beta \in \mathbb{R}^{n+1} \right\} = \min \left\{ \sum_{i=1}^{m} (y_i - f(\beta, x_i))^2 \mid \beta \in \mathbb{R}^{n+1} \right\}, \tag{6}$$

where $||\cdot||$ is the Euclidean norm that defines the objective function of the optimization problem. It can be shown, see [7], that the optimal solution to the latter problem (6) can be found as a solution β to the normal equation

$$X^\top X\beta = X^\top y, \tag{7}$$

where $X^\top$ is the transpose matrix of X.

This introductory part shortly reviewed the problem of the least squares polynomial approximation and introduced notation that is used in the paper. This more demanding problem is described in the next two subsections. First a very special problem with two boundary values is considered. Then a more general problem with different linear bounds is introduced, cf. [4].

2.1 Simple Problem of Least Squares Curve Fitting with Constraints

Now a bounded data set (1) is considered. In particular let

$$D \subset [0, 1] \times [0, 1] \tag{8}$$

and instead of the problem (6) the optimization problem with constraints

$$\min \left\{ \sum_{i=1}^{m} (y_i - f(\beta, x_i))^2 \mid \beta \in \mathbb{R}^{n+1} \wedge f(\beta, 0) = 0 \wedge f(\beta, 1) = 1 \right\}, \tag{9}$$

where f is a polynomial of degree n given by (3), can be solved. To make sense of the given problem it is supposed that the degree n of polynomial f is at least 2, i.e. $n \geq 2$, otherwise a linear function given by two functional values is directly defined.

2.2 General Problem of Least Squares Method with Linear Constraints

A more general description of least squares methods with equality constraints can be introduced. Given matrices $X \in \mathbb{R}^{m \times (n+1)}$, $C \in \mathbb{R}^{p \times (n+1)}$ and vectors $y \in \mathbb{R}^m$, $d \in \mathbb{R}^p$ let

$$\mathscr{M} = \{\beta \in \mathbb{R}^{n+1} \mid C\beta = d\} \tag{10}$$

be a set of constraints. The least squares problem with linear equality constraints is the optimization problem

$$\min\{||y-X\beta||^2 |\ \beta \in \mathscr{M}\}, \tag{11}$$

where $||\cdot||$ is the Euclidean norm that defines the objective function. To illustrate the introduced concepts, we will use a representation of the simple problem from subsection 2.1 now. Let X be as in (5) and

$$C = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{pmatrix}, d = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

then the problem (9) is defined. This fact can be directly checked with the help of relations (12) and (13).

3 Results and Discussion

Two possible algorithms to solve optimization problems (9) and (10),(11) will be introduced here. Both solutions lead to variants of the normal equation, as will be shown now.

3.1 Solution to Simple Problem of Least Squares Curve Fitting with Constraints

The problem (9) will be gradually converted into the standard least squares problem. At first notice that

$$f(\beta,0) = \beta_0$$

and

$$f(\beta,1) = \beta_n + \dots + \beta_2 + \beta_1 + \beta_0.$$

Since it is assumed that two constraints given in (9) are valid it is possible to write

$$\beta_0 = 0 \tag{12}$$
$$\beta_n = 1 - \beta_1 - \beta_2 - \dots - \beta_{n-1} \tag{13}$$

After these facts are used and they are substituted into the system of linear equations (4) the following equations is obtained

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_{n-1} x_i^{n-1} + (1 - \beta_1 - \beta_2 - \dots - \beta_{n-1}) x_i^n, \quad i \in \{1,2,\dots,m\}. \tag{14}$$

The latter system can be simplified and rewritten in the following form

$$y_i - x_i^n = (x_i - x_i^n)\beta_1 + (x_i^2 - x_i^n)\beta_2 + \dots + (x_i^{n-1} - x_i^n)\beta_{n-1}, \quad i \in \{1,2,\dots,m\}. \tag{15}$$

To simplify next consideration it is useful to introduce a more condensed matrix form of the system (15). If the matrix $X^* \in \mathbb{R}^{m\times(n-1)}$ is defined as $X^*_{ij} = x_i^j - x_i^n$, where $i \in \{1,2,\dots,m\}$, $j \in \{1,2,\dots,n-1\}$ and the components of the vector $y^* \in \mathbb{R}^m$ are defined as $y_i^* = y_i - x_i^n$, then the linear system (15) can be written in matrix form

$$X^*\beta^* = y^*,$$

where $\beta^* = (\beta_1, \beta_2, \dots, \beta_{n-1})^\top \in \mathbb{R}^{n-1}$. Notice that the matrix X^* and the vector y^* can be also found with the help of the following matrices $W \in \mathbb{R}^{m\times n}$, $T \in \mathbb{R}^{n\times(n-1)}$ and the vector $e_n \in \mathbb{R}^{n\times 1}$ where

$$W = \begin{pmatrix} x_1 & x_1^2 & \dots & x_1^{n-1} & x_1^n \\ x_1 & x_1^2 & \dots & x_1^{n-1} & x_1^n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_m & x_m^2 & \dots & x_m^{n-1} & x_m^n \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix}, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{16}$$

Then $X^* = WT$ and $y^* = y - We_n$.

Now the least squares solution $\widehat{\beta^*}$ to the problem (15) can be considered as the solution to the optimization problem

$$\min\left\{||y^* - X^*\beta^*||^2 |\ \beta^* \in \mathbb{R}^{n-1}\right\}. \tag{17}$$

This solution can be found as the solution to the normal equation

$$(X^*)^\top X^* \beta^* = (X^*)^\top y^*. \tag{18}$$

Finally, if (12), (13) and (18) are considered together the least squares solution to the problem (9) with constraints can be found as the vector

$$\widehat{\beta} = \begin{pmatrix} 0 \\ \widehat{\beta^*} \\ 1 - \widehat{\beta_1^*} - \widehat{\beta_2^*} - \cdots - \widehat{\beta^*}_{n-1} \end{pmatrix} \in \mathbb{R}^{n+1}. \tag{19}$$

All steps of computations can be now summarized in Algorithm 1.

Algorithm 1: Least squares polynomial approximation with two boundary values, problem (9).

Input : x, y, n (independent vector x, dependent vector y, degree n of polynomial f)
Output: $\widehat{\beta}$ (vector of optimal coefficients of polynomial f)

1 Use vector x, relation (16) and set up matrix W
2 Use n, relation (16) and set up matrix T
3 Use n, relation (16) and set up vector e_n
4 $X^* \leftarrow WT$
5 $y^* \leftarrow y - We_n$
6 Solve equation $(X^*)^\top X^* \beta^* = (X^*)^\top y$ for β^*
7 $\widehat{\beta}_0 \leftarrow 0$,
8 $(\widehat{\beta}_1, \ldots, \widehat{\beta}_{n-1})^\top \leftarrow \beta^*$
9 $\widehat{\beta}_n \leftarrow 1 - (\widehat{\beta}_1 + \cdots + \widehat{\beta}_{n-1})$

3.2 Solution to General Problem

The problem (11) can be solved by Lagrange multiplier method, which is explained e.g. in [4], [1]. First we define Lagrangian

$$L(\beta, \lambda) = \frac{1}{2}||y - X\beta||^2 - \lambda^\top (d - C\beta), \tag{20}$$

where $\lambda \in \mathbb{R}^{p \times 1}$ is a vector of Lagrange multipliers. The scalar $1/2$ before the objective function was used to simplify computations of the optimal solution to the problem (11) and moreover it does not change its value. Also the sign minus before the vector of Lagrange multipliers simplified the resulting equation. According to definition $||y - X\beta||^2 = (y - X\beta)^\top (y - X\beta)$, which is a form that can be used for finding the first derivative of Lagrangian (20). If details of derivative computations are omitted the first order conditions of the optimization problem (10),(11) can be written as the following system of linear equations

$$\frac{\partial L}{\partial \beta}(\widehat{\beta}, \widehat{\lambda}) = -y^\top X + \widehat{\beta}^\top X^\top X + \widehat{\lambda}^\top C = 0, \quad \frac{\partial L}{\partial \lambda}(\widehat{\beta}, \widehat{\lambda}) = -d + C\widehat{\beta} = 0. \tag{21}$$

The first part of the given system of linear equations can be better written with transpose matrices as

$$X^\top X \widehat{\beta} + C^\top \widehat{\lambda} = X^\top y.$$

The second part of the given system of linear equation states that $\widehat{\beta}$ satisfies the equality constraints (or feasibility conditions) given by (10). The system (21) of linear equations for optimal solution $(\widehat{\beta}, \widehat{\lambda})$ can be then rewritten into the following matrix equation with $(n+1)+p$ unknowns

$$\begin{pmatrix} X^\top X & C^\top \\ C & 0 \end{pmatrix} \cdot \begin{pmatrix} \widehat{\beta} \\ \widehat{\lambda} \end{pmatrix} = \begin{pmatrix} X^\top y \\ d \end{pmatrix}, \tag{22}$$

where $0 \in \mathbb{R}^{p \times p}$ is zero matrix with zero elements. This resulting equation can be considered as the normal equation for the constrained least squares problem (10) and (11). All steps of computations can be finally summarized in Algorithm 2.

Algorithm 2: Least squares polynomial approximation with linear constraints, problem (10),(11).

Input : x, y, n, C, d
(independent vector x, dependent vector y, degree n of polynomial f, constraints C,d)

Output: $\widehat{\beta}$ (vector of optimal coefficients of polynomial f)

1 Use vector x, degree n and set up Vandermonde matrix X

2 Set up the matrix

$$A \leftarrow \begin{pmatrix} X^\top X & C^\top \\ C & 0 \end{pmatrix}$$

3 Set up the vector

$$\delta \leftarrow \begin{pmatrix} X^\top y \\ d \end{pmatrix}$$

4 Use (22) and solve the equation $A\gamma = \delta$ for γ

5 $\widehat{\beta} \leftarrow (\gamma_1, \gamma_2, \ldots, \gamma_{n+1})^\top$

4 Application - Approximation of Lorenz Curve

It can often happen that elements with a decisive share in the total sum of the investigated quantity appear in the analyzed problem. For instance a few business organizations dominate a decisive part of the market, a few manufacturers control a critical part of a certain product, a few firms dominate the consumer goods market. Another example is the distribution of household income, where several wealthy individuals receive a large proportion of all household income in a given economy. We are talking about concentration of sales (production, consumption, income) in these cases.

An illustrative idea of concentration can be obtained using a special concentration graph, the so-called Lorenz curve (LC), [9] and [10]. This curve also serves as a tool for describing inequality of distribution of household income in economics, for more details see [8]. Briefly LC is a graphical representation of the accumulated series of ascending (or descending) ordered relative data of a given quantity. In the case of income inequality, LC represents the dependence of the cumulative relative rate of total household income on the cumulative relative number of households that are ranked according to the size of the income. More precisely the Lorenz curve L for cumulative income distribution F with finite value of mean μ is

$$L(x) = \frac{1}{\mu} \int_0^x F^{-1}(t)dt, \quad x \in [0,1], \tag{23}$$

where $F^{-1}(t) = \inf\{z|\ F(z) \geq t\}$ is the left inverse of F or quantile function of F, cf. [5]. By the definition LC passes through the points with coordinates $(0,0)$ and $(1,1)$. In this section the best polynomial approximation of LC for the Czech Republic distribution of household income in 2017 is found.

The appropriate data were extracted from Eurostat, in particular from the table [2], and are shortly summarized in Table 1.

Year/Quintile	1st	2nd	3rd	4th	5th
2017	10.3%	14.7%	17.9%	22.2%	35.0%

Table 1 Distribution of the Czech Republic household income by quantiles, source Eurostat, [2].

Now (23) can be used and data set for LC, similar to (1), can be directly found. The resulting coordinates of LC are introduced in Table 2.

x	0.2	0.4	0.6	0.8	1
$y = L(x)$	0.1030	0.2498	0.4286	0.6503	1.0000

Table 2 Lorenz curve for the Czech Republic household income by quantiles, source: own computations.

Approximation of LC was calculated with both the Algorithm 1 and the Algorithm 2 and identical results were

found. The optimal degree $\widehat{n}$ of the polynomial (3) was found experimentally as the minimum of length of vector of residuals for different values of degree within the finite sets $\{2,3,4,\ldots,10\}$. It was found that $\widehat{n}=4$ and the length of corresponding vector of residuals is $||y-X\widehat{\beta}|| = 2.5\cdot 10^{-14}$. The optimal coefficients of polynomial approximation of LC for the Czech republic in 2017 are summarized in Table 3.

Year/Polynomial Coefficients	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$
2017	0	0.2901	1.4455	-1.9974	1.2618

Table 3 Estimation of the polynomial that approximate the Lorenz curve, source: own computations in Matlab.

5 Conclusion

The article has shown one way to formulate and solve problems in which it is reasonable to look for a polynomial approximation of a functional model using the least squares method with possible boundary conditions. First, a simple problem with two boundary values was introduced. Then a more general problem with linear constraints was considered. Different variants of normal equations (18) and (22) were derived. The introduced problem can be useful in certain economic problems. Here the problem of the polynomial approximation of the Lorenz curve for the Czech Republic in 2017 was solved. Both derived normal equations were used to find the approximation of the Lorenz curve and the numerical results that were found were the same.

There are some unfinished tasks. First, it is necessary to test a possible way how to solve normal equation effectively (22). The algorithm of QR decomposition of a matrix was used in algorithm 2 in line 4. SVD decomposition of this matrix can be considered and tested as the alternative method. Secondly, once the functional model of the Lorenz curve is known, it would be possible to determine directly the value of the Gini coefficient. Thirdly, it is necessary to find more demanding applications of introduced methods in the field of economics.

Acknowledgements

Support of the Specific research project of the Faculty of Informatics and Management of University of Hradec Králové is kindly acknowledged.

References

[1] Bazaraa, M.S., Sherali H.D. and Shetty C.M. (2006). *Nonlinear Programming, Theory and Algorithms*, New Jersey: John Willey & Sons, Inc.

[2] Erostat. *European Union Statistics on Income and Living conditions*. Retrieved April 19, 2019, from https://ec.europa.eu/eurostat/web/income-and-living-conditions/data/database.

[3] Fronckova, K., Prazak, P. (2019). Kalman Filter and Time Series. In Jedlicka, P., Maresova, P. & Soukal, I. (Eds.), *Hradec Economic Days, 9(1)* (pp. 225–234). Hradec Kralove: University of Hradec Kralove.

[4] Gander W., Gander M and J., Kwok F. (2014). *Scientific Computing, An Introduction using Maple and MATLAB*, Heidelberg: Springer.

[5] Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037-1039.

[6] Gilli M., Maringer D. and Schumann E. (2012). *Numerical Methods and Optimization in Finance*, Amsterdam: Elsewier.

[7] Meyer, C.D. (2000). *Matrix Analysis and Applied Linear Algebra*, Philadelphia: SIAM.

[8] Lando, T., Stanickova, M., Franek, J. (2018). Parametric Families for Lorenz Curve: An Analysis of Income Distribution in European Contries. *Ekonomicka revue - Central European Review of Economic Issues*, 21(2), 52–60.

[9] Lorenz, M. O. (1905). Methods for measuring the concentration of wealth. *American Statistical Association* 9, 209–219.

[10] Tureckova K., Kotlanova E. (2015). Method of Non-weighted Average Absolute Deviation in Context of Measuring Income Inequality by Gini Coefficient: case study of Visegrad group countries Article Title. In Martincik, D., Ircingova, J. & Janecek, P. (Eds.), *33rd International Conference Mathematical Methods in Economics* (pp. 835–840). Cheb: University of West Bohemia.

[11] Wooldbridge J.M. (2018). *Introductory Econometrics, A Modern Approach*, Boston: Cengage.

Second Order Optimality in Markov and Semi-Markov Decision Processes

Karel Sladký[1]

Abstract. Semi-Markov decision processes can be considered as an extension of discrete- and continuous-time Markov reward models. Unfortunately, traditional optimality criteria as long-run average reward per time may be quite insufficient to characterize the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect variability-risk features of the problem. Perhaps the best known approaches stem from the classical work of Markowitz on mean-variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance. Such approach has been already studied for very special classes of semi-Markov decision processes, in particular, for Markov decision processes in discrete- and continuous-time setting. In this note these approaches are summarized and possible extensions to the wider class of semi-Markov decision processes is discussed. Attention is mostly restricted to uncontrolled models in which the chain is aperiodic and contains a single class of recurrent states. In this case growth rate of total reward and the variance is again asymptotically linear in time and is independent of the starting state.

Keywords: Semi-Markov processes with rewards, discrete- and continuous-time Markov reward chains, average reward and variance over time, risk-sensitive optimality, policy iterations.

JEL classification: C44, C61
AMS classification: 90C40, 60J10

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize robustness of the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect stability and variability-risk features of the problem. Hence robustness and risk control are also important issues in practical applications. As well known one of the common and popular risk measure is the variance which is often used to characterize the stability of the system. Perhaps the best known approaches stem from the classical work of Markowitz (cf. [6]) on mean variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance. Higher moments and variance of cumulative rewards in Markov reward chains have been primarily studied for discrete time models. Research in this direction has been initiated in Benito [1], Jaquette [4], Mandl [5] and Sobel [15].

In the paper Van Dijk and Sladký [14] results for the discrete-time case are extended to continuous-time Markov reward chains. As the essential step is an expression for the variance of the undiscounted cumulative reward and its asymptotic behavior. In this note, for the sake of simplicity, the presentation is restricted to the *un*controlled case, the implication for the controlled case is only briefly be referred to. For additional results on the limiting average variance for continuous-time models let us mention the paper by Prieto-Rumeau and Hernández-Lerma [7] along with the monograph [2]. Similar results are also reported in Guo et al [3]. Since no transition rewards are considered, the obtained formula for the limiting variance is a special case of more complicated results reported in [14].

The article is structured as follows. Section 2 contains notations and summary of basic facts on discrete- and continuous-time Markov reward chains and their extensions to semi-Markov reward processes. The heart of the paper are sections 3 and 4. Second order optimality for Markov reward chains is discussed in Section 3, i.e. formulas for total expected reward and for the corresponding variances of total reward are derived for discrete- and continuous-time models. The analysis is limited to finite horizon case, however, the obtained results can be also used for long run discounted (or transient) models. Extensions of presented results to semi-Markov reward processes is contained in Section 4. Conclusions are made in Section 5.

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, sladky@utia.cas.cz

2 Notations and Preliminaries

Semi-Markov processes present an extension of Markov processes considered in discrete- and continuous-time setting. Considering Markov models with rewards we can summarize the following facts.

In the discrete-time case, we consider Markov decision chain $X^{\mathrm{d}} = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$, and finite set $\mathcal{A}_i = \{1, 2, \ldots, K_i\}$ of possible decisions (actions) in state $i \in \mathcal{I}$. Supposing that in state $i \in \mathcal{I}$ action $a \in \mathcal{A}_i$ is selected, then state j is reached in the next transition with a given probability $p_{ij}(a)$ and one-stage transition reward r_{ij} will be accrued to such transition.

In the continuous-time setting, the development of the considered Markov decision process $X^{\mathrm{c}} = \{X(t), t \geq 0\}$ (with finite state space $\mathcal{I}$) over time is governed by the transition rates $q(j|i, a)$, for $i, j \in \mathcal{I}$, depending on the selected action $a \in \mathcal{A}_i$. For $j \neq i$ $q(j|i, a)$ is the transition rate from state i into state j, $q(i|i, a) = \sum_{j \in \mathcal{I}, j \neq i} q(j|i, a)$ is the transition rate out of state i. Recall that on entering state i the process stays in state i for a random time that is exponentially distributed with parameter $q(i, a) = -q(i|i, a)$ and the next jump to state j occurs with probability $p_{ij}(a) = q(j|i, a)/q(i, a)$. As concerns reward rates, $r(i)$ denotes the rate earned in state $i \in \mathcal{I}$, and $r(i, j)$ is the transition rate accrued to a transition from state i to state j.

A (Markovian) policy controlling the decision process is given by either a sequence of decision at every time point (discrete-time case) or as a piecewise constant right continuous function of time (continuous-time case). In particular, for discrete-time models policy controlling the chain, $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$ for every $n = 0, 1, 2, \ldots$, and $f_i^n \in \mathcal{A}_i$ is the decision (or action) taken at the nth transition if the chain X^{d} is in state i.

We denote by $P(f) = [p_{ij}(f_i)]$ the $N \times N$ transition matrix of the chain X^{d}. Obviously, the row sums along with the spectral radius of $P(f)$ are equal to one. Transition probability matrix $\tilde{P}(f)$ is called *transient* if the spectral radius of $\tilde{P}(f)$ is less than unity, i.e. it at least some row sums of $\tilde{P}(f)$ are less than one. Then $\lim_{n\to\infty}[\tilde{P}(f)]^n = 0$. Observe that if $P(f)$ is stochastic and $\alpha \in (0, 1)$ then $\tilde{P}(f) := \alpha P(f)$ is transient, however, if $\tilde{P}(f)$ is transient it may happen that some row sums may be even greater than unity.

Policy which takes at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary; $P(f)$ is transition probability matrix with elements $p_{ij}(f_i)$. Recall that the limiting matrix $P^*(f) = \lim_{m\to\infty} m^{-1} \sum_{n=0}^{m-1} P^n(f)$ exists; in case that the chain is aperiodic even $P^*(f) = \lim_{n\to\infty} (P(f))^n$. In particular, if $P(f)$ is *unichain* (i.e. $P(f)$ contains a single class of recurrent states) the rows of $P^*(f)$, denoted $p^*(f_i)$, are identical. Obviously, $r_i(f_i) = \sum_{j=1}^{N} p_{ij}(f_i) r_{ij}$ is the expected one-stage reward obtained in state $i \in \mathcal{I}$ and $r(f)$ denotes the corresponding N-dimensional column vector of one-stage rewards. Then $v(f) := [P(f)]^n \cdot r(f)$ is the (column) vector of rewards accrued after n transitions, its ith entry $v_i(f)$ denotes expectation of the reward if the process X^{d} starts in state i.

Similarly, for the continuous-time case policy controlling the chain, $\pi = f(t)$, is a piecewise constant, right continuous vector function where $f(t) \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$, and $f_i(t) \in \mathcal{A}_i$ is the decision (or action) taken at time t if the process $X(t)$ is in state i. Since π is piecewise constant, for each π we can identify the time points $0 < t_1 < t_2 \ldots < t_i < \ldots$ at which the policy switches; we denote by $f^i \in \mathcal{F}$ the decision rule taken in the time interval $(t_{i-1}, t_i]$. Policy which takes at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary.

Let for $f \in \mathcal{F}$ $Q(f) = [q_{ij}(f_i)]$ be an $N \times N$ matrix whose ijth element $q_{ij}(f_i) = q(j|i, f_i)$ for $i \neq j$ and for the iith element we set $q_{ii}(f_i) = -q(i|i, f_i)$ (recall that the row sums of a transition rate matrix $Q(f)$ are equal to null). The sojourn time of the considered process X^{c} in state $i \in \mathcal{I}$ is exponentially distributed with parameter $q(i|i, f_i)$. Hence the expected value of the reward obtained in state $i \in \mathcal{I}$ equals $r_i(f_i) = [q(i|i, f_i)]^{-1} r(i) + \sum_{j \in \mathcal{I}, j \neq i} q(j|i, f_i)\, r(i, j)$ and $r(f) = [r_i(f)]$ is the (column) vector of reward rates at time t. Recall that 0 is an eigenvalue of $Q(f)$, and the real part of any eigenvalue of $Q(f)$ is non-positive. Similarly to discrete-time model $\tilde{Q}(f)$ is transient if the real part of any eigenvalue of $Q(f)$ is negative.

The above two models can be unified and generalized by introducing semi-Markov reward processes. To this end, we shall define semi-Markov reward processes as follows.

Consider a controlled semi-Markov reward process $Y = \{Y(t), t \geq 0\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ along with the embedded Markov chain $X^{\mathrm{d}} = \{X_n, n = 0, 1, \ldots\}$. We assume that X^{d} is unichain for any stationary policy. The development of the process $Y(t)$ over time is the following:

At time $t = 0$ if $Y(0) = i$ the decision maker selects decision from a finite set $\mathcal{A}_i = \{1, 2, \ldots, K_i\}$ of possible decisions (actions) in state $i \in \mathcal{I}$. Then state j is reached in the next transition with a given probability $p_{ij}(a)$ after random time $\eta_i(a)$. Let $F_i(a, \tau)$ be a non-lattice distribution function representing the conditional probability $\mathrm{P}(\eta_i \leq \tau)$. We assume that for $\ell = 1, 2$ and any $i, j = 1, \ldots, N$, $0 < d_i^{(\ell)}(a) = \int_0^\infty \tau^\ell \,\mathrm{d}F_i(a, \tau) < \infty$. Finally, one-stage transition reward $r(i, j) > 0$ will be accrued to such transition and reward rate $r(i)$ per unit of

time incurred in state i is earned.

A (Markovian) policy controlling the semi-Markov process Y, $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$ for every $n = 0, 1, 2, \ldots$, and $f^n_i \in \mathcal{A}_i$ is the decision (or action) taken at the nth transition if the embedded Markov chain $X^{\rm d}$ is in state i. Let π^k be a sequence of decision vectors starting at the k-th transition, hence $\pi = (f^0, f^1, \ldots f^{k-1}, \pi^k)$.

Let ξ_n be the cumulative random reward obtained in the n first transitions of the considered embedded Markov chain $X^{\rm d}$ and $\xi(t)$ denotes the (random) reward earned up to time t, i.e.

$$\xi_n = \sum_{k=0}^{n-1} [r(X_k) \cdot \eta_{X_k, X_{k+1}} + r(X_k, X_{k+1})], \qquad \xi(t) := \left[\int_0^t r(Y(s)) {\rm d}s \; + \sum_{k=0}^{N(t)} r(Y(\tau_k^-), Y(\tau_k^+)) \right] \tag{1}$$

with $Y(s)$, denoting the state of the system at time s, $Y(\tau_k^-)$ and $Y(\tau_k^+)$ the state just prior and after the kth jump, $N(t)$ the number of jumps up to time t.

Obviously, discrete-time Markov reward chain is a very special case of semi-Markov reward process where all holding times are non-random and equal to one, and one-stage rewards depend only on the labels of the consecutive two states. Moreover, continuous-time Markov reward chains can be considered as a very specific case of semi-Markov reward processes where holding times are exponentially distributed. In particular, considering continuous-time Markov reward process with transition rates $q(j|i, a)$, reward rates $r(i)$ in state i and rewards per transition $r(i, j)$, the process can be treated as a semi-Markov process with transition probabilities $p_{ij}(a) = q(j|i, a)/q(i, a)$ and exponentially distributed holding times with parameter $q(i, a)$.

It is well-known that the long-run average reward of the considered semi-Markov process Y can be calculated using the embedded Markov chain $X^{\rm d}$. In particular, if stationary policy $\pi \sim (f)$ is followed, on recalling that $X^{\rm d}$ is unichain the limiting matrix $P^*(f)$ has identical rows, i.e. $p^*_j(f)$ is the jth entry of each row of $P^*(f)$. Moreover, for the long run models also the fraction of time spent by the semi-Markov process Y in state i can be easily calculated (see e.g. [8, 9]). Then the average reward per unit of time, say $\bar{g}(f)$, generated by the semi-Markov process Y is independent of the starting state and can be calculated as

$$\bar{g}(f) = \sum_{j \in \mathcal{I}} \bar{p}^*_j(f) \cdot r_j(f), \text{ where } \bar{p}^*_i(f) = \frac{p^*_i(f) \cdot d_i(f)}{\sum_{j \in \mathcal{I}} p^*_{j\ell}(f) \cdot d_j(f)}, \; r_j(f) = d_j(f) \cdot r(j) + \sum_{\ell \in \mathcal{I}} p_{j\ell}(f_i) \cdot r(j, \ell). \tag{2}$$

Similarly it is possible to extend the presented discrete-time Markov reward chain model to a more general model of semi-Markov reward processes. To this end, let $\eta_i(a)$ be the random time spent in state i with expectation $d_i(a)$ if action a is chosen. Obviously, it suffices to add in each state $i \in \mathcal{I}$ to the one-stage rewards r_{ij} the following term $r(i) \cdot \eta_i(a)$ representing additional reward earned during random stay of the process X in state i. Hence the expected value and the second moment of the total reward earned in state i are $r(i) \cdot d^{(1)}_i(a) + \sum_{i \in \mathcal{I}} p_{ij}(a) \cdot r(i, j)$ and $\sum_{i \in \mathcal{I}} p_{ij}(a) \cdot \mathsf{E}\,[r(i) \cdot \eta_i(a) + r_{ij}]^2$ respectively.

3 Second Order Optimality in Markov Reward Chains

Considering discrete-time models, let $\xi_n(\pi) = \sum_{k=0}^{n-1} r_{X_k, X_{k+1}}$ be the stream of rewards received in the n next transitions of the considered Markov chain X if policy $\pi = (f^n)$ is followed. Supposing that $X_0 = i$, on taking expectation we get for the first and second moments of $\xi_n(\pi)$

$$v^{(1)}_i(\pi, n) := \mathsf{E}^{\,\pi}_i(\xi_n(\pi)) = \mathsf{E}^{\,\pi}_i \sum_{k=0}^{n-1} r_{X_k, X_{k+1}}, \qquad v^{(2)}_i(\pi, n) := \mathsf{E}^{\,\pi}_i(\xi_n(\pi))^2 = \mathsf{E}^{\,\pi}_i (\sum_{k=0}^{n-1} r_{X_k, X_{k+1}})^2.$$

If policy $\pi \sim (f)$ is stationary, the process $X^{\rm d}$ is time homogeneous and for $m < n$ we write for the generated random reward $\xi_n = \xi_m + \xi_{n-m}$ (here we delete the symbol π and tacitly assume that $\mathrm{P}(X_m = j)$ and ξ_{n-m} starts in state j). Hence $[\xi_n]^2 = [\xi_m]^2 + [\xi_{n-m}]^2 + 2 \cdot \xi_m \cdot \xi_{n-m}$. Then for $n > m$ we can conclude that

$$\mathsf{E}^{\,\pi}_i[\xi_n] \;=\; \mathsf{E}^{\,\pi}_i[\xi_m] + \mathsf{E}^{\,\pi}_i \big\{ \sum_{j \in \mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}^{\,\pi}_j[\xi_{n-m}] \big\}. \tag{3}$$

$$\mathsf{E}^{\,\pi}_i[\xi_n]^2 \;=\; \mathsf{E}^{\,\pi}_i[\xi_m]^2 + \mathsf{E}^{\,\pi}_i \big\{ \sum_{j \in \mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}^{\,\pi}_j[\xi_{n-m}]^2 \big\} + 2 \cdot \mathsf{E}^{\,\pi}_i[\xi_m] \sum_{j \in \mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}^{\,\pi}_j[\xi_{n-m}]. \tag{4}$$

In particular, from (3), (4) we conclude for $m = 1$

$$v_i^{(1)}(f, n+1) = r_i^{(1)}(f_i) + \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot v_j^{(1)}(f, n) \tag{5}$$

$$v_i^{(2)}(f, n+1) = r_i^{(2)}(f_i) + 2 \cdot \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot r_{ij} \cdot v_j^{(1)}(f, n) + \sum_{j \in \mathcal{I}} p_{ij}(f_i)\, v_j^{(2)}(f, n) \tag{6}$$

where $r_i^{(1)}(f_i) := \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot r_{ij}$, $r_i^{(2)}(f_i) := [\sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot [r_{ij}]^2$.

Since the variance $\sigma_i^{(2)}(f, n) = v_i^{(2)}(f, n) - [v_i^{(1)}(f, n)]^2$ from (5),(6) we get

$$\begin{aligned} \sigma_i^{(2)}(f, n+1) &= r_i^{(2)}(f_i) + \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot \sigma_j^{(2)}(f, n) + 2 \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot r_{ij} \cdot v_j^{(1)}(f, n) \\ &\quad - [v_i^{(1)}(f, n+1)]^2 + \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot [v_j^{(1)}(f, n)]^2 \qquad (7) \\ &= \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot [r_{ij} + v_j^{(1)}(f, n)]^2 - [v_i^{(1)}(f, n+1)]^2 + \sum_{j \in \mathcal{I}} p_{ij}(f_i) \cdot \sigma_j^{(2)}(f, n). \qquad (8) \end{aligned}$$

Using matrix notations (cf. [12, 13]) equations (5),(6),(20) can be written as:

$$v^{(1)}(f, n+1) = r^{(1)}(f) + P(f) \cdot v^{(1)}(f, n) \tag{9}$$

$$v^{(2)}(f, n+1) = r^{(2)}(f) + 2 \cdot P(f) \circ R \cdot v^{(1)}(f, n) + P(f) \cdot v^{(2)}(f, n) \tag{10}$$

$$\begin{aligned} \sigma^{(2)}(f, n+1) &= r^{(2)}(f) + P(f) \cdot \sigma^{(2)}(f, n) + 2 \cdot P(f) \circ R \cdot v^{(1)}(f, n) \\ &\quad - [v^{(1)}(f, n+1)]^2 + P(f) \cdot [v^{(1)}(f, n)]^2 \qquad (11) \end{aligned}$$

where $R = [r_{ij}]$ is an $N \times N$-matrix, and $r^{(2)}(f) = [r_i^{(2)}(f_i)]$, $v^{(2)}(f, n) = [v_i^{(2)}(f, n)]$, $v^{(1)}(f, n) = [(v_i^{(1)}(f, n)]$, $\sigma^{(2)}(f, n) = [\sigma_i^{(2)}(f, n)]$ are column vectors. The symbol $\circ$ is used for Hadamard (entrywise) product of matrices. Observe that $r^{(1)}(f) = (P(f) \circ R) \cdot e$, $r^{(2)}(f) = [P(f) \circ (R \circ R)] \cdot e$ (e is reserved for unit column vector).

Similarly, considering Markov reward chains in continuous time the expected reward $v_i(t, \pi)$ can be considered as the first moment of the random variable $\xi(t)$ if the starting state $X(0) = i$ policy $\pi = f(t)$ is followed. Similarly, the corresponding second moment and variance are given by $v_i^{(2)}(t, \pi) := \mathsf{E}_i^{\pi}[\xi(t)]^2$, $\sigma_i^{(2)}(t, \pi) := v_i^{(2)}(t, \pi) - [v_i(t, \pi)]^2$.

Considering stationary policy $\pi \sim (f)$, let $\xi(t + \Delta) = \xi(\Delta) + \xi^{(\Delta, t+\Delta)}$ where $\xi^{(\Delta, t+\Delta)}$ is reserved for the total (random) reward obtained in the time interval $[\Delta, t + \Delta)$. Then $\xi(\Delta) + \xi^{(\Delta, t+\Delta)}$ and $[\xi(t + \Delta)]^2 = [\xi(\Delta)]^2 + [\xi^{(\Delta, t+\Delta)}]^2 + 2[\xi(\Delta)][\xi^{(\Delta, t+\Delta)}]$ and hence

$$\mathsf{E}_i^{\pi}[\xi(t + \Delta)] = \mathsf{E}_i^{\pi}[\xi(\Delta)] + \mathsf{E}_i^{\pi}[\xi^{(\Delta, t+\Delta)}] \tag{12}$$

$$\mathsf{E}_i^{\pi}[\xi(t + \Delta)]^2 = \mathsf{E}_i^{\pi}[\xi(\Delta)]^2 + \mathsf{E}_i^{\pi}[\xi^{(\Delta, t+\Delta)}]^2 + 2 \cdot \mathsf{E}_i^{\pi}[\xi(\Delta)][\xi^{(\Delta, t+\Delta)}] \tag{13}$$

Then, by using that $P(\Delta, f) = I + \Delta Q(f) + o(\Delta^2)$ and that the probability for more than one transition in time Δ is of order Δ^2, for Δ tending to zero we obtain

$$\begin{aligned} \frac{\mathrm{d}v_i(t, f)}{\mathrm{d}t} &= r(i) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot r(i, j) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot [v_j(t, f) - v_i(t, f)] \\ &= r_i(f_i) + \sum_{j \in \mathcal{I}} q_{ij}(f_i) \cdot v_j(t, f) \qquad (14) \end{aligned}$$

$$\frac{\mathrm{d}v_i^{(2)}(t, f)}{\mathrm{d}t} = 2 \cdot r(i) \cdot v_i(t, f) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \left\{ [r(i, j)]^2 + 2 \cdot r(i, j) \cdot v_j(t, f) \right\} + \sum_{j \in \mathcal{I}} q_{ij}(f_i) \cdot v_j^{(2)}(t, f) \tag{15}$$

By $\sigma_i^{(2)}(t,f) = v_i^{(2)}(t,f) - [v_i(t,f)]^2$ we thus obtain:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sigma_i^{(2)}(t,f) &= \frac{\mathrm{d}}{\mathrm{d}t}v_i^{(2)}(t,f) - 2\cdot v_i(t,f)\frac{\mathrm{d}}{\mathrm{d}t}v_i(t,f)\\ &= 2\cdot r(i)\cdot v_i(t,f) + \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\left\{[r(i,j)]^2 + 2\cdot r(i,j)\cdot v_j(t,f)\right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i)\cdot v_j^{(2)}(t,f)\\ &\quad -2\cdot v_i(t,f)\cdot r(i)(f_i) + \sum_{j\in\mathcal{I}} q_{ij}(f_i)\cdot v_j(t,f) \qquad (16)\end{aligned}$$

Using matrix notations equations (14),(15) can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}v(t,f) = r(f) + Q(f)\cdot v(t,f), \qquad \frac{\mathrm{d}}{\mathrm{d}t}v^{(2)}(t,f) = r^{(2)}(t,f) + Q(f)\cdot v^{(2)}(t,f). \qquad (17)$$

where $r(f) = [r_i(f)]$, $r^{(2)}(t,f) = [r_i^{(2)}(t,f)]$, $v(t,f) = [v_i(t,f)]$, $v^{(2)}(t,f) = [v_i^{(2)}(t,f)]$, are column vectors with elements $r_i(f_i) = r(i) + \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)r(i,j)$, $r_i^{(2)}(t,f) = 2r(i)v_i(t,f) + \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\{[r(i,j)]^2 + 2r(i,j)v_j(t,f)\}$.

Similarly after some algebra (16) can be also written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma^{(2)}(t,f) = r^{(2\sigma)}(t,f) + Q(f)\sigma^{(2)}(t,f) \qquad \text{where}$$

$\sigma^{(2)}(t,f) = [\sigma_i^{(2)}(t,f)]$, $r^{(2\sigma)}(t,f) = [r_i^{(2\sigma)}(t,f)]$, $r_i^{(2\sigma)}(t,f) = \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)[r(i,j) + v_j(t,f) - v_i(t,f)]^2$

4 Mean Reward Variance and Semi-Markov Processes

In this section we extend reported results concerning Markov reward chains to semi-Markov processes. To this end, we focus attention on the embedded Markov chain X^{d} and extend the corresponding formulas presented in Section 3. In particular, we shall assume that if in state $i \in \mathcal{I}$ action a is selected and state j is then reached the one-step (random) reward earned in state i, say $\xi_{i,j}$, is equal to $r(i)\cdot\eta_i(a) + r(i,j)$ and taking the expectation we can conclude that $\mathsf{E}\,\xi_{i,j} = i\sum_{j\in\mathcal{I}} p_{ij}(a)\cdot r(i,j) + r(i)\cdot d_i^{(1)}(a)$.
Unfortunately, $\mathsf{E}\,[\xi_i]^2 = [r(i)]^2\cdot d_i^{(2)}(a) + \sum_{j\in\mathcal{I}} p_{ij}(a)\{2r(i)\cdot d_i^{(1)}(a) + [r(i,j)]^2\}$ and on using this way of reasoning formulas (3)–(6) can be replaced by

$$v_i^{(1)}(f,n+1) = r_i^{(1)}(f_i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot v_j^{(1)}(f,n) \qquad (18)$$

$$v_i^{(2)}(f,n+1) = r_i^{(2)}(f_i) + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot[r(i,j) + r(i)\cdot d_i(f_i)\cdot v_j^{(1)}(f,n)] + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot v_j^{(2)}(f,n) \qquad (19)$$

where $r_i^{(1)}(f_i) := r(i)\cdot d_i^{(1)}(f_i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot r(i,j)$, $r_i^{(2)}(f_i) := [\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot[r(i,j)]^2$.

Since the variance $\sigma_i^{(2)}(f,n) = v_i^{(2)}(f,n) - [v_i^{(1)}(f,n)]^2$ from (5),(6) we get

$$\begin{aligned}\sigma_i^{(2)}(f,n+1) &= r_i^{(2)}(f_i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{(2)}(f,n) + 2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot r(i,j)\cdot v_j^{(1)}(f,n)\\ &\quad -[v_i^{(1)}(f,n+1)]^2 + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot[v_j^{(1)}(f,n)]^2 \qquad (20)\\ &= \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot[r_{ij} + v_j^{(1)}(f,n)]^2 - [v_i^{(1)}(f,n+1)]^2 + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{(2)}(f,n). \qquad (21)\end{aligned}$$

The long-run average variance (independent of the starting state) is

$$\bar{g}(f) = \sum_{j\in\mathcal{I}} \bar{p}_j^*(f)\cdot\bar{r}_j(f), \text{ where } \bar{p}_i^*(f) = \frac{p_i^*(f)\cdot d_i(f)}{\sum_{j\in\mathcal{I}} p_{j\ell}^*(f)\cdot d_j(f)},\ \bar{r}_j(f) = d_j(f)\cdot r(j) + \sum_{\ell\in\mathcal{I}} p_{j\ell}(f_i)\cdot r(j,\ell).$$

5 Conclusions

Solving problem on stochastic dynamic programming the decision maker selects by standard policy or value iteration methods the set of all optimal e.g. maximazing average reward. In the next step the decision maker selects in the class of optimal policies policies according to the to second order optimality criterion.

In this note formula for calculating long-run average variance of unichain semi-Markov reward processes is obtained. This also extend results concerning average variance for discrete- a continuous-time Markov reward chains. In particular, solving problems on stochastic dynamic programming at first the decision maker find by standard policy or value iteration methods the set of all optimal policies. In the next step the decision maker selects in the class of optimal policies policies according to the to second order optimality criterion.

Acknowledgement

This research was supported by the Czech Science Foundation under Grant 18-02739S.

References

[1] Benito, F. (1982). Calculating the variance in Markov processes with random reward. *Trabajos de Estadistica y de Investigacion Operativa* 33, 73–85.

[2] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes: Theory and Applications.* Berlin: Springer.

[3] Guo, X. and Song, X. (2009). Mean-variance criteria for finite continous-time Markov decision processes. *IEEE Trans. Automat. Control* 54, 2151–2157.

[4] Jaquette, S.C. (1973). Markov decision processes with a new optimality criterion: Discrete time. *Ann. Statist.* 1, 496–505.

[5] Mandl, P. (1971). On the variance in controlled Markov chains. *Kybernetika* 7, 1–12.

[6] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7, 77–92.

[7] Prieto-Rumeau, T. and Hernández-Lerma, O. (2009). Variance minimization and the overtaking optimality approach to continuous-time controlled Markov chains. *Math. Method Oper. Res.* 70, 527–540.

[8] Puterman, M.L. (1994). *Markov Decision Processes – Discrete Stochastic Dynamic Programming.* New York: Wiley.

[9] Ross, S.M. (1983). *Introduction to Stochastic Dynamic Programming.* New York: Academic Press.

[10] Sladký K. (2005). On mean reward variance in semi-Markov processes. *Math. Methods Oper. Res.* 62, 387-397.

[11] Sladký, K. (2013). Risk-sensitive and mean variance optimality in Markov decision processes. *Acta Oeconomica Pragensia* 7, 146–161.

[12] Sladký, K. (2017). Second order optimality in Markov decision chains. *Kybernetika* 53, 1086–1099.

[13] Sladký, K. (2018). Risk-sensitive and mean variance optimality in continuous-time Markov decision chains. In: 36th Internat. Conf. Mathem. Methods in Economics (L.Váchová, V.Kratochvíl, eds.), Jindřichův Hradec 2018, pp. 497–512.

[14] Van Dijk, N.M. and Sladký, K. (2006). On total reward variance for continuous-time Markov reward chains. *J. Appl. Probab.*. 43, 1044–1052.

[15] Sobel, M.J. (1982). The variance of discounted Markov decision processes. *J. Appl. Probab.* 19, 794–802.

Critical Mass Task Identification in Projects

Tomáš Šubrt[1], Jan Bartoška[2], Petr Kučera[3]

Abstract. Critical Mass is a term that refers to some project activities which, even though not on the Critical Path, are very risky for in time project completion. Tasks potentially involved in critical mass are generally small, low priority, miscellaneous work, not linked to many others. A difference between activities on the Critical Path and those comprising a Critical Mass is that Critical Path activities must be completed as scheduled, in order to achieve the target completion date. Critical Mass activities generally have plenty of float or slack and can be executed just about any time and in any order. The aim of our paper is an identification of these activities based on quantitative indicators and to propose early warning system for project managers based on quantitative indicators, that such tasks exist and that they need to be financed and monitored. Our approach is an extension and modification of multi-criteria approach to activities criticalness and their expansion to the turnaround aspect, the degree of routine and isolation degree of the task. As the tasks potentially involved in critical mass are inadequately resourced, we will also pay attention to resource utilization indicators and factors limiting resource motivation to work on critical mass tasks.

Keywords: Project management, Critical Mass, multiple criteria decision making, task criticalness

JEL Classification: C44
AMS Classification: 90C29, 91B32

1 Introduction

Modern project management works with a variety of methods that help it identify tasks and / or tasks resources that potentially threaten the in time completion of the project, respectively its completion at all. From the 1950s to the 1980s, there were almost exclusively the critical path methods, such as CPM, MPM, PERT, GERT, whose common philosophy was to identify tasks that delay or prolong a project's completion date due to delay or extension of tasks. Based on the results that these methods provide the project managers can draw their attention to critical tasks that need to be stressed in terms of conditions for their implementation, resource allocation and sufficient budgeting. For other tasks, they define potential threats, but the tasks highly uncritical (with big slacks) are by definition omitted. Newer methods from the last quarter of the 20th century are more focused on the human factor in project activities completing. Multi-resource tasks, or sequences of activities with a single assigned resource, are seen as a threatening project [7]. Thus, the Critical Chain method and its derivatives seek to implement project management practices that eliminate human factors (Parkinson's law, multitasking, student syndrome) and, with the use of post-limitation postulates [4], warn the project manager of potential project threats or failures. As in the case of critical path procedures, the tasks of the project manager are, by definition, large, long, financially and resource-intensive. Other methods assessing the importance - criticality of activities are heuristic, based, for example, based on the so-called potential criticality of activities [3], measuring their different indicators with the maximum or aggregate characteristics of the project in terms of cost, resources, duration, connectivity, distance from origin and draw their importance or potential criticality for the project. In general, the higher the value of criticality, the more it needs to focus on the activity. In the rather managerial area, the effort to manage the key stages of the project, respectively the project as a whole, is focused on monitoring the progress and involvement of people in processes and controls. Such new approaches are often Agile-based methods [8].

Critical Mass-based procedures, however, are based on a completely different premise - the manager must not forget about small and unimportant tasks, which can gather and endanger the project just as big, critical once.

[1] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, subrt@pef.czu.cz.
[2] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[3] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, kucera@pef.czu.cz.

2 Materials and methods

2.1 Turnaround and Critical Mass

The first thoughts that small and insignificant activities may endanger the project due date arose at the beginning of 21st century, when the Turnaround "project" Management theory, emerges in the context of hurricane Katrina in 2005. After huge destruction in New Orleans it was necessary to prove some new and very progressive project management techniques to rebuilt the city as fast as possible. But overall the goal was not completely clear. It is necessary to work quickly, using tasks or projects analogy and routine operation. On the other hand, due to big uncertainties in projects, there had to be frequent checks of the progress, often partial objectives and schedules were redefined or redesigned. In such context of Turnaround, there may be tasks that, by concentrating into a critical mass, can endanger the project. By generalization, it turns out that such tasks are not typical only for turnaround but also for classic projects.

Critical Mass is a term that refers to the many activities which, even though not on the Critical Path, have a potential to extend beyond the target completion date. Such activities typically not exceed ten percent of the estimated man-hours for a turnaround or project. Even though critical path activities may be on schedule, insufficient manpower could force the bulk of other activities to be scheduled late. The jobs involved in critical mass are generally smaller, low priority, miscellaneous work. One interesting difference between activities on the Critical Path and those comprising a Critical Mass is that Critical Path activities must be executed (completed) as scheduled, in order to achieve the target completion date. Critical Mass activities, on the other hand, generally have plenty of float (slack time) and can be executed just about any time (in any order) [7].

2.2 Task Criticalness Potential

The suggested approach to the evaluation of a task in relation to a project success is based on the crisp quantitative approach. This approach suggests providing an overall evaluation of the task criticalness without soft knowledge of the task nature. The estimation of overall criticalness potential of the project tasks is based on the multiple attributes decision-making method MAUT using five criteria: topological location, duration, slack, cost, and work and corresponding criticalness indicators which are defined as utility function for each criterion. The real value of all these criteria have to be normalized into the interval $\langle 0,1 \rangle$ so that the worst value of the criteria is transformed to value 1. Using such transformation we obtain task criticality indicators. After the aggregation of partial indicators the task with the highest value of overall criticalness potential is the most threatened task. The project manager can set his/her own preferences which can be used in this aggregation. Five indicators of the task criticalness are defined: (i) topological criticalness, (ii) time criticalness, (iii) slack criticalness, (iv) cost criticalness, and (v) work criticalness. These indicators are subsequently used as an input data for the multiple attribute evaluation of the task criticalness potential [1], [2].

The principle of calculating most of the indicators is similar in this approach, and so, for example, a procedure for cost criticality indicator can be listed.

This indicator is defined from the perspective of minimizing project cost. We can assume the task with low cost has a smaller impact on the total cost of the project than an expensive one. This indicator expresses the relative cost of each project activities

$$cc_i = \frac{c_i - \min\limits_{k=1,2,..N} c_k}{\max\limits_{k=1,2,..N} c_k - \min\limits_{k=1,2,..N} c_k} \tag{1}$$

where cc_i is the cost criticalness of the task i, c_i, c_k resp., are the cost of tasks i, k, resp., N is the number of the tasks in the project. The cost criticalness indicator of the task transforms the task cost so that the higher value of this indicator shows higher criticalness.

Indicators of the task criticalness are used as an input for multiple criteria decision-making models. Values of all indicators are within interval $\langle 0,1 \rangle$ and used as an input for multiple attribute analysis model. Usually it is used a Multiplicative and an Additive multiple criteria approach for evaluation of the criticalness potential of tasks

[5]. These models are based on the idea that the higher value of indicators leads to higher criticalness potential of a task. [2].

Although our approach is based on the concept of criticality, it focuses on low criticality values and expands them with other indicators typical of critical mass threats.

3 Results and Discussion

The starting point of our approach is the concept of criticality, enriched by other indicators following the specific features of the Critical Mass Member Candidate (CMMC) tasks, which may endanger the project itself or its crucial parts. As mentioned above, tasks of this type are often isolated, can take place at any time, and many successors are not dependent on their completion. The first criterion for the assessment is the continuity criterion, in other words the degree of isolation of the task. The number of incident edges with a given node representing the project activity is crucial for its determination (we assume that the project is formalized by an Activity On Node network graph). The incoming edges do not play an important role; there is no control mechanism for initiating the next activity for predecessors. Therefore, the continuity criterion can be derived from the output degree of the current node. If output degree is greater than 2, sufficient successors can be assumed to control task completion, in the opposite site, there is lack of control of task completion.

The continuity criterion was set as

$$ci_i = \frac{deg_b i}{2} \tag{2}$$

where ci_i is criterion of task continuity, deg_b i is output degree of the node i . This criterion is relevant if and only if less or equal to 1.

Another criterion showing potential task neglect is the criteria of resource diversity and resource intensity. Again, it is considered that CMMC activities are not resource intensive or resource variable. The analysis shows neglect threating, if no more than two resources and/or no more than two assignment units appears. Therefore, it can be derived the

Criterion of resource variability

$$crv_i = \frac{k_i}{2} \tag{3}$$

where crv_i is criterion of resource variability, k_i is number of resource types assigned to task i . This criterion is relevant if and only if less or equal to 1.

Criterion of resource intensity

$$cri_i = \frac{\sum_{k=1}^{s} r_i^k}{2} \tag{4}$$

where cri_i is criterion of resource intensity, r_i^k is number of resource units assigned to task i, k is number of resource types assigned to task i, s is number of resource types assigned to task i (maximum 2). This criterion is relevant if and only if less or equal to 1.

These two criteria can be combined to the resource neglect criterion as a product of previous two

$$cr_i = crv_i . cri_i \tag{5}$$

This criterion is relevant if and only if less or equal to 1.

Next criteria for potential CMMC are directly derived from the potential of criticality, with the difference that the aggregate indicator of the entire project (total duration, budget, total work effort) is always taken as the normalization basis (in fractionator). The reason for this is to eliminate the influence of dominant activity on this indicator when using the difference between the most demanding and the least demanding activity (see formula (1)). Therefore, we can define three additional criteria:

Task duration neglect criterion

$$ct_i = \frac{t_i}{T} \tag{6}$$

where ct_i is task duration neglect criterion, t_i is duration of task i, T is total duration of the project.

Task cost neglect criterion

$$cc_i = \frac{c_i}{C} \tag{7}$$

where cc_i is task cost neglect criterion, c_i is cost of task i, C is project budget at completion.

Task work neglect criterion

$$cw_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{8}$$

where cw_i is task work neglect criterion, w_i is amount of work of the task i, N is the number of tasks in the project

A special position among the neglect criteria is the neglect criterion of the slack, based on the idea that CMMC are activities that can be performed practically any time during project timespan, thus having a large total slack relative to the duration of the project. This criterion would be the only one of a maximization type, so it is already to be converted into minimization one.

Task slack neglect criterion

$$cs_i = 1 - \frac{s_i}{T} \tag{9}$$

where cs_i is task slack neglect criterion, s_i is total slack of task i, T is total duration of the project.

The overall risk of the project failure due to CMMC task can be assessed by multiple attributes decision-making method in accordance with the criticalness concept assessment procedures. The most suitable method seems to be the Simple weighted additive method, which results in the values of partial utility functions. In our case, the most neglect threatened task would be an activity whose value of the utility function would be minimal. Therefore, the authors introduce the more suitable term Task Neglect Potential, expressed using the value of partial utility function as 1-utility function value. The setting of individual criteria weights for this concept may seem problematic. According to practice, the slack neglect criterion and duration neglect criteria are probably the most important, but this cannot be generalized. It is clear from the definition of individual criteria that all of them are minimizing with values within the interval <0,1>. It is thus possible to use a directly convex linear combination of weights and criteria values, i.e. all criteria evaluations are combined using the formula

$$\Gamma F_i = 1 - (u_1 ci_i + u_2 cr_i + u_3 ct_i + u_4 cc_i + u_5 cw_i + u_6 cs_i) \;\; i = 1,2,..,N \tag{10}$$

where ΓF_i is global evaluation of the task neglect potential, u_1, u_2, u_3, u_4, u_5 are the weights of partial criteria

The routine of tasks is considered an essential feature for determining Critical Mass [7], i.e. situation when tasks are repeated several times during a project progress in the same or analogous form. This phenomenon is quite difficult to capture. In practice, these tasks can be named identically (Substation Test 1, Substation Test 2….), or assigned the same resources (revision technicians). The authors are trying to determine another way of assessment them. It is based on the similarity (equality) of the criteria. By placing criteria values into a matrix, their potential similarity and thus the risk of routine can be evaluated.

$$TRM = \begin{pmatrix} ci_1 & crv_1 & cri_1 & ct_1 & cc_1 & cw_1 & cs_1 \\ ci_2 & crv_2 & cri_2 & ct_2 & cc_2 & cw_2 & cs_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ ci_m & crv_m & cri_m & ct_m & cc_m & cs_m & cs_m \end{pmatrix}$$

where *TRM* is task potential routine matrix, *m* is a number of CMMC tasks

If at least $TRM_i \approx TRM_j \approx TRM_k \;\; i \neq j \neq k$ it can be assumed that this is a recurring (routine) activity, again potentially threatened by neglect. The more lines in the TRM are the same, the more routine activities can occur.

4 Case Study

The Critical Mass approach provides relevant results for larger and large projects, because the possibility of neglecting the timely execution of any task is small for small projects. However, let's assume the following project whose indicators, as well as the time schedule (Gantt Chart), are created using MS Project (see Figure 1).

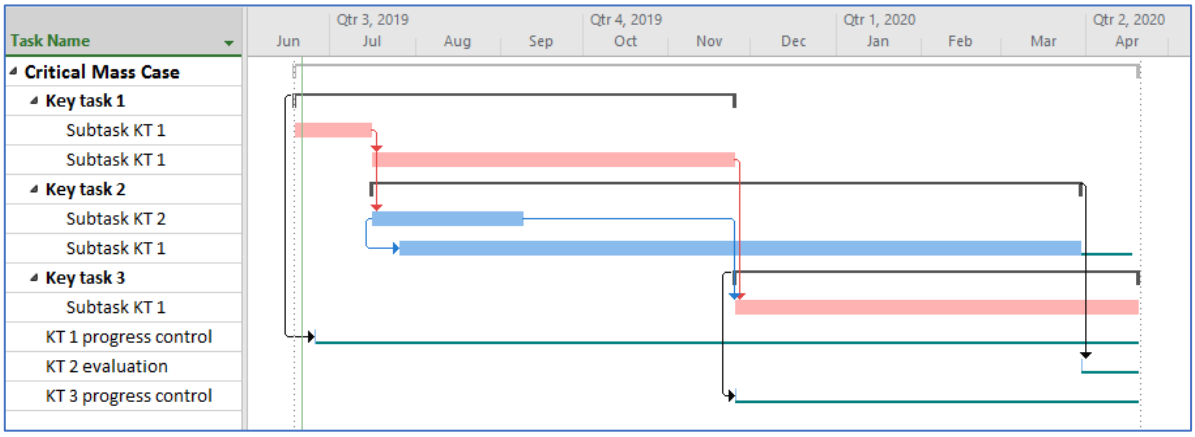


Figure 1 Case study: Critical Mass Project Gantt Chart

All the tasks have assigned resources and work and all the resources are rated. Summary durations, costs and work parameters together with partial neglect criterions are presented in the following table (Table 1). For summary tasks the neglect criteria have no meaning. If continuity and resource neglect criteria are bigger than one, the neglect potential can still be calculated, but the final value of utility function (neglect potential) is thus distorted – values with gray backgrounds in cells.

Evaluation of the last column in Table 1 – Task Neglect Potential – shows, that the highest risk of neglecting has “KT 1 progress control”, followed by “KT 3 progress control”. In case of bigger project, such type of tasks can potentially form a Critical Mass threating the project as a whole.

Task Name	Duration in dys	Total slack in dys	Cost in CZK	Work in hrs	Number of resources	Number of resource units	Number of successors	cii	crvi	crii	cri	cti	cci	cwi	csi	TASK NEGLECT POTENCIAL
Criteria weights								*0,125*	*0,125*	*0,125*	*0,125*	*0,125*	*0,125*	*0,125*	*0,125*	
Critical Mass Case	**218,88**	**0**	**13351080,00**	**8876**				0	0	0	0	1	1	1	1	
Key task 1	**113,88**	**0**	**3292160,00**	**1822**				0	0	0	0	0,520285	0,246584	0,205273	1	
Subtask KT 1	20	0	288000,00	320	2	2	1	0,5	1	1	1	0,091374	0,021571	0,036052	1	**0,4189**
Subtask KT 1	93,88	0	3004160,00	1502	1	2	2	1	1	1	1	0,428911	0,225013	0,16922	1	**0,2721**
Key task 2	**183,88**	**14,95**	**6446480,00**	**4532**			1									
Subtask KT 2	38,88	14,95	466560,00	311	1	1	2	1	0,5	0,5	0,25	0,177632	0,034945	0,035038	0,931698	**0,5713**
Subtask KT 1	175,88	14,95	5979920,00	4221	2	3	0	0	1,5	1,5	2,25	0,803545	0,447898	0,475552	0,931698	**0,0114**
Key task 3	**105**	**0**	**3612000,00**	**2520**			0									
Subtask KT 1	105	0	3612000,00	2520	2	3	0	0	1,5	1,5	2,25	0,479715	0,27054	0,283912	1	**0,0895**
KT 1 progress control	0,1	213,78	2000,00	0,8	1	1	0	0	0,5	0,5	0,25	0,000457	0,00015	9,01E-05	0,0233	**0,8408**
KT 2 evaluation	0,05	14,95	1000,00	0,4	1	1	0	0	0,5	0,5	0,25	0,000228	7,49E-05	4,51E-05	0,931698	**0,7272**
KT 3 progress control	0,07	104,93	1400,00	0,56	1	1	0	0	0,5	0,5	0,25	0,00032	0,000105	6,31E-05	0,520605	**0,7786**

Table 1 Case study: Task parameters and indicators

5 Conclusion

To miss something unimportant in managerial practice sometimes happens. However, projects are sequences of activities in which each one has its own irreplaceable place. Failure of any of them may have fatal consequences. Classical quantitative approaches in project management possess a large number of hard and soft tools to detect activities, or resources risky to meeting project goals. Usually, using such methods, tasks having a limited time frame for their completion, are diverse or costly and otherwise challenging, are marked as project threatening. Procedure proposed in this article allows to find and analyze, in terms of the Critical Mass method, another activities which, on the basis of their indicators (neglect criteria), could also endanger the whole project or significantly reduce the quality of its indicators. Our approach should help project managers to have a more wide and diverse set of tools performing project management more and more complex without drawing distinctions between complex and complicated [6].

References

[1] Brožová, H., Rydval, J. (2017). Modelling synergy of the complexity and criticalness factors in the Project management. In *Proceedings of the 35th International Conference on Mathematical Methods in Economics 13.09.2017, Hradec Králové.* Hradec Králové: Univerzita Hradec Králové, 2017. s. 84-89.

[2] Brožová, H., Bartoška, J., Šubrt, T., Rydval, J. (2016). Task Criticalness Potential: A Multiple Criteria Approach to Project Management. *Kybernetika*, 2016, vol. 52/ 4, pp. 558-574. ISSN: 0023-5954.

[3] Brožová, H., Dvořák, P., Šubrt, T. (2014). Criticalness and the threat of project tasks. In *Proceedings of the 14th International Conference on Project Management and Scheduling 30.03.2014, Munich, Germany.* Munich, Germany: TUM School of Management , pp. 52-56.

[4] Goldratt, E.M. (1997). *Critical Chain*. The North River Press, MA

[5] Mota, C. M. M., de Almeida, A.T. (2011). A multicriteria decision model for assigning priority classes to activities in project management. *Annals of Operations Research*, *199*, 361–37.

[6] Whitty, S. J., Maylor, H. (2009). And then came Complex Project Management (revised). *International Journal of Project Management*, *27*, 304–310.

[7] Turnaround Project Management Primer - Justification for Managing Turnarounds. *InterPlan Systems - STO Estimating, Planning, Scheduling and Management Software* [online]. Copyright © 2002 [cit. 13.05.2019]. Available on: https://www.interplansystems.com/turnaround-project-management-primer/

[8] 12 Principles Behind the Agile Manifesto | Agile Alliance. *Agile Alliance* [online]. Copyright © 2019 Agile Alliance. All Rights Reserved. [cit. 13.05.2019]. Available on: https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto/

Mean-Risk Optimization Problem via Scalarization, Stochastic Dominance, Empirical Estimates

Vlasta Kaňková [1]

Abstract. Many economic and financial situations depend simultaneously on a random element and on a decision parameter. Mostly it is possible to influence the above mentioned situation by an optimization model depending on a probability measure.We focus on a special case of one–stage two–objective stochastic "Mean–Risk problem". Of course to determine optimal solution simultaneously with respect to the both criteria is mostly impossible. Consequently, it is necessary to employ some approaches. A few of them are known (from the literature), however two of them are very important; the first of them is based on a scalarizing technique and the second one is based on the stochastic dominance. The first approach has been suggested (in a special case) by Markowitz, the second approach is based on the second order stochastic dominance. The last approach corresponds (under some assumptions) to partial order in the set of the utility functions.
The aim of the contribution is to deal with the both main approaches mentioned above. First, we repeat their properties and further we try to suggest possibility to improve the both values simultaneously with respect to the both criteria. However, we focus mainly on the case when probability characteristics has to be estimated on the data base.

Keywords: Two–objective stochastic optimization problems, scalarization, stochastic dominance, empirical estimates

JEL classification: C44
AMS classification: 90C15, 90C29, 60E15

1 Introduction

Let $(\Omega, \mathcal{S}, P)$ be a probability space, $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ an s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$, $F := F_\xi$ the distribution function of ξ, P_F, and Z_F the probability measure and the support corresponding to F, respectively. Let, moreover, $g_0 : \mathcal{R}^n \times \mathcal{R}^s \to \mathcal{R}^1$ be real-valued function, $X \subset \mathcal{R}^n$ a nonempty "deterministic" set; E_F, $\rho := \rho_F$ denote the operator of mathematical expectation and the operator of risk measure corresponding to the distribution function F. To introduce mean-risk model let for $x \in X$ there exist finite $\mathsf{E}_F g_0(x, \xi)$, $\rho_F g_0(x, \xi)$. An objective optimization mean–risk problem can be defined as two–objective problem in the following form:

$$\text{Find} \quad \max \mathsf{E}_F g_0(x, \xi), \quad \min \rho_F(g_0(x, \xi)) \quad \text{s.t.} \quad x \in X. \tag{1}$$

Evidently, to optimize simultaneously both objectives is mostly impossible. Different approaches are known from the literature. We recall, at first, three of them; consequently we define new problems:
a.

$$\text{Find} \quad \max \mathsf{E}_F g_0(x, \xi) \qquad \text{s.t.} \qquad \rho_F(g_0(x, \xi)) \le \nu_1, \; x \in X, \tag{2}$$

b.

$$\text{Find} \quad \min \rho_F(g_0(x, \xi)) \qquad \text{s.t.} \qquad \mathsf{E}_F g_0(x, \xi) \ge \nu_2, \; x \in X, \tag{3}$$

c. Markowitz approach [10]

$$\text{Find} \quad \min\{(1 - \lambda)\mathsf{E}_F[-g_0(x, \xi)] + \lambda \rho_F(g_0(x, \xi))\} \quad \text{s.t.} \quad x \in X; \quad \lambda \in \langle 0, 1 \rangle. \tag{4}$$

(ν_1, ν_2 are suitable constants.)

Evidently, the properties of the problems (2), (3), (4) depend on the probability measure P_F, on the properties of the function $g_0(x, \xi)$ and on the risk measure $\rho_F(\cdot)$. We recall a few well–known risk measures ρ_F. To this end we set $U = g_0(x, \xi)$; they are:

[1]The Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, kankova@utia.cas.cz

1. variance – $\rho_F(U)(:= \mathrm{var}(U)) = \mathsf{E}_F[U - \mathsf{E}_F U]^2$,
2. absolute semi–deviation – $\rho_F(U)\,(:= \bar{\delta}(U)) = \mathsf{E}_F[\max\,(\mathsf{E}_F[U] - U),\, 0]$,
3. the standard semi–deviation – $\rho_F(U)\,(:= (\delta(U)) = (\mathsf{E}_F[(\max(\mathsf{E}_F[U] - U,\, 0))^2])^{1/2}$,
4. $\rho_F(U)\,(:=$ Average Value-at-Risk) = $AV@R_\alpha(U)$ for some fixed $\alpha \in [0,\, 1]$.
   (For the definition of $AV@R_\alpha(U)$ see, e.g., [4].)

Moreover, there exists a relationship between the Mean-Risk model (1), Markowitz approach and the stochastic dominance approach. We employ it for the second order stochastic dominance [4]. However, to introduce the corresponding definition we have to recall, first, a definition of the second order stochastic dominance. To this end let $Y(:= Y(\xi(\omega)))$, $V(:= V(\xi(\omega)))$ be random variables defined on $(\Omega, \mathcal{S}, P)$. If there exist finite $\mathsf{E}_F V(\xi)$, $\mathsf{E}_F Y(\xi)$ and if

$$F^2_{Y(\xi)}(u) = \int\limits_{-\infty}^{u} F_{Y(\xi)}(z)dz, \quad F^2_{V(\xi)}(u) = \int\limits_{-\infty}^{u} F_{V(\xi)}(z)dz, \quad u \in \mathcal{R}^1,$$

then $Y(\xi)$ dominates in second order $V(\xi)$ $(Y(\xi) \succeq_2 V(\xi))$ if

$$F^2_{Y(\xi)}(u) \le F^2_{V(\xi)}(u) \quad \text{for every} \quad u \in \mathcal{R}^1.$$

Definition 1. [4] The mean–risk model (1) is called consistent with the second order stochastic dominance ($\succeq_2$) if for every $x \in X$ and $y \in X$,

$$g_0(x,\, \xi) \succeq_2 g_0(y,\, \xi) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x,\, \xi) \ge \mathsf{E}_F g_0(y,\, \xi) \quad \text{and} \quad \rho_F(g_0(x,\, \xi)) \le \rho_F(g_0(y,\, \xi)). \tag{5}$$

According to the definition of the second order stochastic dominance $g_0(x,\, \xi) \succeq_2 g_0(y,\, \xi)$ means that

$$F^2_{g_0(x,\,\xi)}(u) = \int\limits_{-\infty}^{u} F_{g_0(x,\,\xi)}(z)dz \le F^2_{g_0(y,\,\xi)}(u) = \int\limits_{-\infty}^{u} F_{g_0(y,\,\xi)}(z)dz \quad \text{for every} \quad u \in \mathcal{R}^1. \tag{6}$$

Employing the results [11], the stochastic second order dominance (6) can be rewritten in a more friendly form:

$$\mathsf{E}_F(u - g_0(x,\xi))^+ \le \mathsf{E}_F(u - g_0(y,\xi))^+ \quad \text{for every} \quad u \in \mathcal{R}^1. \tag{7}$$

Evidently to analyze the above mentioned approaches we can employ already known results for deterministic and stochastic optimization. In the next part we repeat some of them. However, before it we try to introduce an organization of the paper. A brief survey of the corresponding definitions and auxiliary assertions is given in Section 2. Section 3 is devoted to Markowitz approach, results determined on the base of the second order stochastic dominance can be found in Section 4. The contribution is closed by Conclusion (Section 5).

2 Some Definitions and Auxiliary Assertions

Setting $f_1(x) = -\mathsf{E}_F g_0(x,\, \xi), \quad f_2(x) = \rho_F(g_0(x,\, \xi))$ we can see that problem (1) is a problem of two-objective deterministic optimization. Consequently it is possible for their investigation to employ the results achieved for multi-objective deterministic problems.

2.1 Deterministic Multi-Objective Problems

To recall suitable results obtained for deterministic problems, let $f_i(x)$, $i = 1, \ldots, l$ be real–valued functions defined on $\mathcal{R}^n$; $\mathcal{K} \subset \mathcal{R}^n$ be a nonempty set. The multi–objective deterministic optimization problem can be defined by:

$$\text{Find} \quad \min f_i(x),\, i = 1, \ldots, l \quad \text{subject to} \quad x \in \mathcal{K}. \tag{8}$$

Definition 2. The vector x^* is an efficient solution of the problem (8) if and only if there exists no $x \in \mathcal{K}$ such that $f_i(x) \le f_i(x^*)$ for $i = 1, \ldots, l$ and such that for at least one i_0 one has $f_{i_0}(x) < f_{i_0}(x^*)$.

Definition 3. The vector x^* is a properly efficient solution of the multi–objective optimization problem (8) if and only if it is efficient and if there exists a scalar $M > 0$ such that for each i and each $x \in \mathcal{K}$ satisfying $f_i(x) < f_i(x^*)$ there exists at least one j such that $f_j(x^*) < f_j(x)$ and

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \le M. \tag{9}$$

Proposition 1. [3] *Let $\mathcal{K} \subset \mathcal{R}^n$ be a nonempty convex set and let $f_i(x)$, $i = 1, \ldots, l$ be convex functions on $\mathcal{K}$. Then $x^0 \in \mathcal{K}$ is a properly efficient solution of the problem (8) if and only if x^0 is optimal in*

$$\text{Find} \quad \min_{x \in \mathcal{K}} \sum_{i=1}^{l} \lambda_i f_i(x) \quad \text{for some} \quad \lambda_1, \ldots, \lambda_l > 0, \quad \sum_{i=1}^{l} \lambda_i = 1.$$

A relationship between efficient and properly efficient points is introduced, e.g., in [2] or [3]. We summarize it in the following Remark.

Remark 1. Let $f(x) = (f_1(x), \ldots, f_l(x))$, $x \in \mathcal{K}$; $\mathcal{K}^{eff}$, $\mathcal{K}^{peff}$ be sets of efficient and properly efficient points of the problem (8). If $\mathcal{K}$ is a convex set, $f_i(x)$, $i = 1, \ldots, l$ are convex functions on $\mathcal{K}$, then

$$f(\mathcal{K}^{peff}) \subset f(\mathcal{K}^{eff}) \subset \bar{f}(\mathcal{K}^{peff}), \quad (\bar{f}(\mathcal{K}^{peff} \quad \text{denotes the closure set of} \quad f(\mathcal{K}^{peff})).$$

Remark 2. Evidently setting $l = 2$, $\mathcal{K} = X$ and $f_1(x) = -\mathsf{E}_F g_0(x, \xi), \quad f_2(x) = \rho_F(g_0(x, \xi))$ we can see that problem (1) corresponds to the deterministic problem (8).

Further, we recall the definition of strongly convex function.

Definition 4. Let $h(x)$ be a real-valued function defined on a nonempty convex set $\mathcal{K} \subset \mathcal{R}^n$. $h(x)$ is strongly convex function with a parameter $\rho' > 0$ if

$$h(\lambda x^1 + (1-\lambda)x^2) \leq \lambda h(x^1) + (1-\lambda)h(x^2) - \lambda(1-\lambda)\rho' \|x^1 - x^2\|_2^2$$

$$\text{for very} \quad x^1, x^2 \in \mathcal{K}, \ \lambda \in \langle 0, 1\rangle, \quad (\|\cdot\|_2 := \|\cdot\|_2^n \quad \text{denotes the Euclidean norm in} \quad \mathcal{R}^n).$$

2.2 Stochastic Optimization Problems and Empirical Estimates

Let $\mathcal{P}(\mathcal{R}^s)$ denote the set of all (Borel) probability measures on $\mathcal{R}^s$, $\mathcal{M}_1^1(\mathcal{R}^s)$ be defined by the relation:

$$\mathcal{M}_1^1(\mathcal{R}^s) := \left\{ \nu \in \mathcal{P}(\mathcal{R}^s) : \int_{\mathcal{R}^s} \|z\|_1 d\nu(z) < \infty \right\}, \quad \|\cdot\|_1^s := \|\cdot\|_1 \text{ denotes } \mathcal{L}_1 \text{ norm in } \mathcal{R}^s. \tag{10}$$

Let, further, $g : \mathcal{R}^n \times \mathcal{R}^s \to \mathcal{R}^1$ be real-valued function such that there exists a finite $\mathsf{E}_F g(x, \xi)$ for every $x \in X$. We introduce a system of assumptions:

A.0 $g(x, z)$ is for $x \in X$ a Lipschitz function of $z \in \mathcal{R}^s$ with the Lipschitz constant (corresponding to the $\mathcal{L}_1$ norm) not depending on x,

A.1 $g(x, z)$ is either a uniformly continuous function on $X \times \mathcal{R}^s$ or there exists $\varepsilon > 0$ such that $g(x, z)$ is a convex function on $X(\varepsilon)$ and bounded on $X(\varepsilon) \times \mathcal{R}^s$ ($X(\varepsilon)$ denotes ε–neighborhood of X),

A.2
- $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F,
- F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^{N}, N = 1, 2, \ldots,$
- $F_i, i = 1, \ldots, s$ denote one-dimensional marginal distribution functions corresponding to F.

Let, further, for the random value $Y := Y(\xi)$ and $\varepsilon \in \mathcal{R}^1$ the sets $X_F^{\varepsilon} := X_F^{\varepsilon}(Y(\xi))$ be defined by

$$X_F^{\varepsilon} = \{x \in X : \mathsf{E}_F(u - g_0(x, \xi))^+ - \mathsf{E}_F(u - Y(\xi))^+ \leq \varepsilon \quad \text{for every} \quad u \in \mathcal{R}^1\}, \quad \varepsilon \in \mathcal{R}^1. \tag{11}$$

If we set $X_F = X_F^0$, then a rather general optimization problem with second order stochastic dominance constraints can be introduce in the following form:

$$\text{Find} \quad \varphi(F, X_F) = \inf\{\mathsf{E}_F g(x, \xi) : x \in X_F\}. \tag{12}$$

Replacing the distribution function F by the empirical one F^N, then we obtain an empirical optimization problem. The following assertion follows from stability results presented in [9].

Proposition 2. [9] *Let X_F be a nonempty compact set, $P_F \in \mathcal{M}_1^1(\mathcal{R}^s)$. Let, moreover, $g_0(x, z)$, $Y(z)$ be for every $x \in X$ Lipschitz functions of $z \in Z_F$ with the Lipschitz constant not depending on $x \in X$. If*

1. *Assumptions A.0, A.1, A.2 are fulfilled,*
2. *$g(x, z)$ is a Lipschitz function on X with the Lipschitz constant not depending on $z \in Z_F$,*
3. *there exists $\varepsilon_0 > 0$ such that X_F^{ε} are nonempty compact sets for every $\varepsilon \in \langle -\varepsilon_0, \varepsilon_0\rangle$ and, moreover, there exists a constant $\hat{C} > 0$ such that*

$$\Delta_n[X_F^{\varepsilon}, \ X_F^{\varepsilon'}] \leq \hat{C}|\varepsilon - \varepsilon'| \quad \text{for} \quad \varepsilon, \varepsilon' \in \langle -\varepsilon_0, \varepsilon_0\rangle,$$

4. there exists a finite first moment of the random vector ξ,

then

$$P\{\omega : |\varphi(F, X_F) - \varphi(F^N, X_{F^N}| \longrightarrow_{N\longrightarrow\infty} 0\} = 1 \tag{13}$$

A crucial assumptions (in Proposition 2) is the existence of a finite first moment of ξ. Consequently, the relation (13) holds also for stable distributions with the parameter stability greater or equal to 1 (for the definition of stable distribution see, e.g., [8]).

3 Mean-Risk Model via Markowitz Approach

Evidently, the following assertion follows from Proposition 1.

Proposition 3. *Let $X \subset \mathcal{R}^n$ be a nonempty convex set, $\mathsf{E}_F[-g_0(x, \xi)]$, $\rho_F(g_0(x, \xi))$ be finite convex functions on X. Then $x^\lambda \in X$ is a properly efficient solution of the problem (1) if and only if x^λ is optimal in the problem*

$$\text{Find} \quad \min_{x\in X}\{(1-\lambda)\mathsf{E}_F[-g_0(x, \xi)] + \lambda\rho_F(g_0(x, \xi))\} \quad \text{for} \quad \lambda \in (0, 1). \tag{14}$$

If we denote by $\bar{X}$ the set of all solutions of the problem (14) for some $\lambda \in (0, 1)$, then according to Remark 1 the closure of the set $\bar{X}$ is equal to set of all efficient points of two-objective problem (1). If, moreover, functions $\mathsf{E}_F[-g_0(x, \xi)]$, $\rho_F(g_0(x, \xi))$ are strongly convex on X with the same parameter ρ' and if $\bar{\mathcal{X}}(F, X)$, $\bar{\mathcal{G}}(F, X)$ are defined by the relation:

$$\begin{aligned}
\bar{\mathcal{X}}(F, X) &= \{x \in X : x \quad \text{is a properly efficient point of the problem (1)}\},\\
\bar{\mathcal{G}}(F, X) &= \{t_1, t_2 : t_1 = \mathsf{E}_F[-g_0(x, \xi)],\ t_2 = \rho_F(g_0(x, \xi)) \quad \text{for some} \quad x \in \bar{\mathcal{X}}(X, F)\},
\end{aligned}$$

then the following assertion follows from [6] and [7].

Theorem 1. *Let P_F, $P_G \in \mathcal{M}_1^1(\mathcal{R}^s)$, X be a nonempty convex compact set, A.2 be fulfilled. If*

- *$g_0(x, z)$ is for every ($x \in X$) a Lipschitz function of $z \in \mathcal{R}^s$ with the Lipschitz constant (corresponding to $\mathcal{L}_1$ norm) not depending on $x \in X$,*
- *there exists a constant $C^1 > 0$ such that*

$$|\rho_F(g_0(x, \xi) - \rho_G(g_0(x, \xi))| \le C^1 \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)|dz_i,$$

- *$\mathsf{E}_F g_0(x, \xi)$, $\rho_F(g_0(x, \xi))$ are strongly convex on X with the same parameter $\rho' > 0$,*

then

$$P\{\omega : \Delta_n[\bar{\mathcal{X}}(F, X) - \bar{\mathcal{X}}(F^N, X)] \to_{N\to\infty} 0\} = 1.$$

(Symbol $\Delta[\cdot, \cdot] := \Delta_n[\cdot, \cdot]$ denotes the Hausdorff distance of two nonempty sets in $\mathcal{R}^n$, for the definition Hausdorff distance see, e.g., [14].)

Remark 3. Conditions under which variance and Avarage-value-at-Risk fulfil the assumptions of Theorem 1 can be found in [7].

4 Mean-Risk Model via Second Order Stochastic Dominance

Let us start with model (1), problems (2), (3) and Definition 1. If we can for suitable ν_1, ν_2 find $x_0 \in X$ such that

$$\mathsf{E}_F g_0(x_0, \xi) \ge \nu_2, \quad \rho_F g_0(x_0, \xi)) \le \nu_1,$$

then setting $Y(\xi) = g_0(x_0, \xi)$ we can define the set $X(x_0)$ by

$$X(x_0) = \{x \in X : \mathsf{E}_F(u - g_0(x, \xi))^+ \le \mathsf{E}_F(u - g_0(x_0, \xi))^+ \quad \text{for every} \quad u \in \mathcal{R}^1\}. \tag{15}$$

In the case when the model (1) is consistent with the second order stochastic dominance and $X(x_0)$ is a nonempty set, then

$$x \in X(x_o) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \ge \mathsf{E}_F g_0(x_0, \xi) \quad \text{and simultaneously} \quad \rho_F(g_0(x, \xi)) \le \rho_F(g_0(x_0, \xi)).$$

Further, evidently, if we can determine $x_1 \in X(x_0)$ such that $\mathsf{E}_F g_0(x_1, \xi) > \mathsf{E}_F g_0(x_0, \xi)$, then setting $Y(\xi) := g_0(x_1, \xi)$, we can define the set $X(x_1)$ by

$$X(x_1) = \{x \in X(x_1) : \mathsf{E}_F(u - g_0(x, \xi))^+ \leq \mathsf{E}_F(u - g_0(x_1, \xi))^+ \quad \text{for every} \quad u \in \mathcal{R}^1\}. \tag{16}$$

Employing Definition 1 we obtain the following relations

$$x \in X(x_1) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \geq \mathsf{E}_F g_1(x_1, \xi) \quad \text{and simultaneously} \quad \rho_F(g_0(x, \xi)) \leq \rho_F(g_0(x_1, \xi)).$$

Further, evidently, if we can determine $x_2 \in X(x_1)$ such that that $\mathsf{E}_F g_0(x_2, \xi) > \mathsf{E}_F g_0(x_1, \xi)$, then we can setting $Y(\xi) := g_0(x_2, \xi)$, employing Definition 1 define the set $X(x_2)$ by

$$X(x_2) = \{x \in X(x_2) : \mathsf{E}_F(u - g_0(x, \xi))^+ \leq \mathsf{E}_F(u - g_0(x_2, \xi))^+ \quad \text{for every} \quad u \in \mathcal{R}^1\} \tag{17}$$

such that

$$x \in X(x_2) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \geq \mathsf{E}_F g_1(x_2, \xi) \quad \text{and simultaneously} \quad \rho_F(g_0(x, \xi)) \leq \rho_F(g_0(x_2, \xi)).$$

Of course, we can continue looking for $x_3, x_4, \ldots$.

Evidently

$$X(x_2) \subset X(x_1) \subset X(x_0),$$

$$x \in X(x_2) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \geq \mathsf{E}_F g_0(x_2, \xi) > \mathsf{E}_F g_0(x_1, \xi) > \mathsf{E}_F g_0(x_0, \xi),$$

$$\rho_F g_0(x, \xi) \leq \rho_F g_0(x_2, \xi) \leq \rho_F g_0(x_1, \xi) \leq \rho_F g_0(x_0, \xi).$$

Of course, theoretically, it is possible to determine already $x_1 \in X(x_0)$ by an optimization problem:

$$\text{Find} \quad \max \mathsf{E}_F g_0(x, \xi) \quad \text{s.t.} \quad x \in X(x_0). \tag{18}$$

However, problem (18) is generally semi-infinite optimization problem for which Slater's condition is not very often fulfilled. Many authors have dealt with this problems, we can recall e.g., [1], [5], [9], [5].

Till now (in this section) we have assumed that the underlying probability measure is known. This assumption is fulfilled in real situations very seldom and the problem has to be analyzed on the data base. We can recall works dealing with this situation see, e.g., [4], [9] and [15].

Remark 4.

- It has been proven in [13] (see also [4])) that the mean-risk problem using Average Value-at Risk at some level α (as risk measure) is consistent with the second order stochastic dominance.
- Of course, it is possible to interchange the positions of $\mathsf{E}_F g_0(x, \xi)$ and $\rho_F(g_0(x, \xi))$ and consequently to try to improve the value of risk measure.

5 Conclusion

We have dealt with a special case of multi-objective stochastic optimization problems. Especially we consider Mean-Risk problem with Markowitz approach and with the approach based on the stochastic second order dominance. It could seem that our results do not cover many real situations. However, to deal with these others cases is beyond of the scope of this contribution.

Acknowledgment This work was supported by the Czech Science Foundation under grant 18-02739S.

References

[1] Dentcheva, D. and Ruszczynski, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization* 14, 548–566.

[2] Ehrgott, M. (2005). *Multicriteria Optimization.* Springer: Berlin.

[3] Geoffrion, A.M. (1968). Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.* 22, 3, 618–630.

[4] Gutjahr, W.J. and Pichler, A. (2013). Stochastic multi–objective optimization: a survey on non–scalarizing methods. *Annals of Operations Research*.

[5] Hu, J., Tito-Homem-de-Mello and Mehrotra, S. (2012). Sample average approximation of stochastic dominance constrined programs. *Math. Program. Ser. A* 133, 171–201.

[6] Kaňková, V. and Houda, M. (2015). Thin and heavy tails in stochastic programming. *Kybernetika* 51, 3, 433–456.

[7] Kaňková, V. (2015). A remark on multi–objective stochastic optimization with strongly convex functions. *CEJOR* 24, 309–333.

[8] Kaňková, V. and Omelchenko, V. (2015). Empirical estimates in stochastic programs with probability and second order dominance constraints. *Acta Mathematica Universitatis Comenianae* 84, 2, 267–281.

[9] Kaňková, V. and Omelchenko, V. (2018). Stochastic optimization problem with second order stochastic dominance constrains via Wasserstein metric. *Kybernetika* 54, 6, 1234–1246.

[10] Markowitz, H.M. (1952). Portfolio selection. *J. of Finance* 6, 77–91.

[11] Ogryczak, W. and Rusczynski, A. (1999). From the stochastic dominance to mean–risk models: semideviations as risk measure. *European J. Oper. Res.* 116, 33–50.

[12] Ogryczak, W. and Rusczynski, A. (2001). On consistency of stochastic dominance and mean–semideviation models. *Math. Program. Ser.B.* 89, 217–232.

[13] Ogryczak, W. and Rusczynski, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13, 1, 60–78.

[14] Rockafellar, R.T. and Wets, R.J-B. (1983). *Variational Analysis.* Springer: Berlin.

[15] Sun, H., Xu, H., Meskarian, R., and Wang, Y.: Exact penalization, level function method and modified cutting plane method for stochastic programs with second order dominance constraints. *SIAM Journal on Optimization* 23, 602–631.

The use of belief function theory in recommendation based on a similarity diffusion

Ladislav Beranek[1], Radim Remes[2]

Abstract. At present, recommendation systems are integral part of recommendation systems in e-business. In these systems, collaborative filtering technology is an important method for assessing user preferences by using user feedback data and is widely used. Diffusion-based recommendation based on diffusion phenomenon is an important method in collaborative filtering recommendation for processes that can be modelled by a bipartite network, e.g., processes which can represent behavior between users and e-shops. Diffusion-based recommendation algorithms calculate the similarities between users and make recommendations only regarding implicit feedback but neglect the benefits of explicit feedback which is formed by texts written by users during feedback process next to evaluation by certain amount of points (implicit feedback). These texts can be however a significant element in recommendation systems. This paper proposes a combined diffusion similarity model for the integration of explicit feedback and implicit feedback based on Dempster-Shafer theory. The experimental results show that the proposed solution gives good results compared to other algorithms.

Keywords: Recommendation system, E-business, Network, Diffusion, Collaborative filtering.

JEL Classification: C63
AMS Classification: 91D30

1 Introduction

The Internet and related business activities are developing rapidly. The amount of information is growing rapidly and the overload of information seems to be a serious problem. Recommendation systems [8] have been recognized as an effective tool to address this problem and play a key role a process when users view internet and seek some information. Recommendation systems play an important role in areas such as recommendation of movies [6], [14], contents [11], e-commerce services [15], and others. Collaborative filtering is a typical and most popular information technology in recommendation systems [17]. Its main idea is to evaluate user preferences using user feedback data. Two kinds of feedback data can be processed, i.e., explicit feedback and implicit feedback. The first, e.g., 1 to 5 points rating, means the level a user likes, while the other, such as clicks or purchases, indicates whether a user likes an item or not [7]. In addition, elements in an explicit feedback matrix can be arbitrary numeric values, while the implicit feedback matrix is a single-value matrix.

Diffusion-based recommendation algorithms are based on network structures that are inspired by diffusion in physical dynamics [13, 16]. These algorithms use often a bipartite network of user items to represent input data, such as evaluation matrices, and bipartite network links indicate data collection behavior between users and items.

This paper proposes a combined model of diffusion similarity. It is designed to include both explicit feedback data and implicit feedback data, inspired by the idea of hybrid diffusion [18], taking into account user levels and items to improve performance in diffusion processes. In doing so, we also use the belief functions theory [10] to express some uncertainty associated with implicit and explicit user ratings and their similarities.

The rest of the post is organized as follows. Section 2 presents the related work of diffusion-based recommendation algorithms. Section 3 proposes a description of the proposed model. The results of the experiments are presented in Chapter 4. Section 5 concludes this work with discussion and future work.

[1] University of South Bohemia/Faculty of Economics, Department Applied Mathematics and Informatics, Studentska 13. Ceske Budejovice, beranek@ef.jcu.cz.
[2] University of South Bohemia/Faculty of Economics, Department Applied Mathematics and Informatics, Studentska 13. Ceske Budejovice, inrem@ef.jcu.cz.

2 Literature background

Collaborative filtering is one of the most widely used approaches in recommendation systems. Recently, diffusion-based recommendation algorithms are beginning to be explored. They are often based on bipartite networks or tripartite networks [18]. Diffusion-based recommendation algorithms are inspired by the diffusion phenomenon in physical dynamics, which simulates the process of allocating resources to bipartite user networks for making recommendations. In particular, some discussions about bipartite user item networks can be found in [21]. Mass diffusion (MD) [21] and heat conduction (HC) [19] are considered pioneers of diffusion-based recommendation algorithms. These two approaches assume that each item collected by the target user has one unit of initial resource and builds a two-step random walk process that redistributes the resource on bipartite networks [29]. The bulk diffusion model distributes the resource based on the degree of each node and focuses on the accuracy of the recommendation, while the heat conduction model redistributes the resource based on the degrees of adjacent nodes of each node to suggest high diversity recommendations. Authors in [5] introduced the similarity index among items into a bulk diffusion model to increase accuracy, and authors in [9] have proposed a preferential dissemination model that takes into account the diversity of user levels. There are also studies on the heat conduction model. Specifically, Liu et al. [19] have designed a biased heat conduction model that reduces the attention of small-scale objects to increase both accuracy and diversity. Some previous studies suggest that reducing the amount of initial resources to high grade items can improve the accuracy of recommendations [1]. Time information plays a crucial role in recommendation systems, some time-oriented approaches bring substantial improvements in accuracy by reducing the impact of outdated data [12].

3 The proposal of a recommendation mechanism

The recommended system can be modeled as a bipartite graph $G(U, V, E)$, whose vertices can be divided into two disjunct and independent sets U ($|U| = m$) and V ($|V| = n$) so that each edge joins the vertex u in U with one v in V. We denote the set of edges as E, where $E \subseteq U \times V$. The set E ($|E| = r$) demonstrates relations between user and product. In our concept, U denotes a set of buyers (users), V denotes a set of products. The neighborhood matrix A is defined as follows:

$$a_{ij} = \begin{cases} 1, the\ i\ user\ have\ bought\ the\ product\ j \\ 0, else \end{cases} \quad (1)$$

The degree k_i of the node i of the set U set represents the number of products purchased by the user i. The degree of k_j of the node j of the set V in the bipartite graph represents the number of users who bought the product j. The purpose of recommendation system is to evaluate user preferences and provide a list of recommendations for the target user. This means that a set of items with the highest referral score will be included in the referral list.

The Mass Diffusion Model (MD) is a successful and popular recommendation algorithm [21] that uses the resource allocation process to make recommendations in a bipartite network. In MD, the target user must be selected first. Thereafter, this user will receive a list of recommended items. The items associated with the target user in the bipartite network then receive the initial source. MD can be described as a two-stage resource allocation process. In step 1, the initial resource on item nodes flows to adjacent user nodes based on the degree of each item, so that the resource on each user node can be calculated as:

$$\phi_{ij} = \sum_{l=1}^{n} \frac{a_{il} a_{jl}}{k_l} \varphi_l \quad (2)$$

User i is the target user and user j is the user who obtains the resource in the first step. Next, φ_l and Φ_{ij} are the initial resources of item l and resource per user j after the first step and a_{il} and a_{jl} are the elements of the adjacency matrix A, k_l is the degree of item 1. The distribution strategy in step 2 is based on the degree of each user. The final source Φ_{ij} on an item r is defined as follows:

$$\phi_{ir} = \sum_{j=1}^{m} \frac{a_{jr}}{k_j} \phi_{ij} \quad (3)$$

where k_j is the degree of a user j.

After two steps of the resource allocation process, the initial resource is redistributed to the items, and then a list of non-selected items by final source on each item can be organized for the target user. Unselected items with the highest source will be placed at the top of the list. An illustration of the resource allocation process of the mass diffusion model is shown in Figure 1.

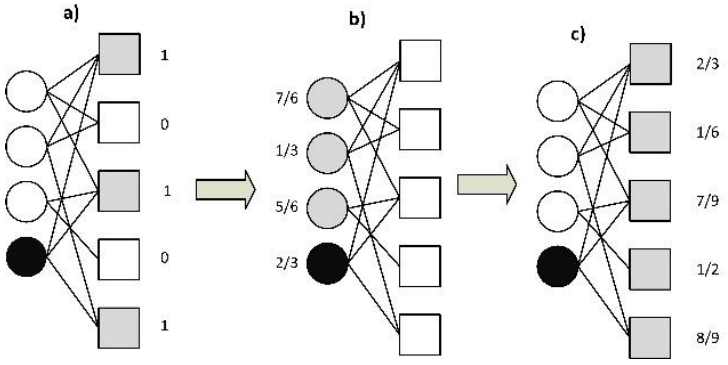


Figure 1 An illustration of mass diffusion model (MD) [21].

Users are represented by circles, and items by squares. The black circle means the target user. Circles and squares with grey color indicate the resources that are distributed on these nodes. Plot (a) is the initial configuration, each item linked to the target user obtains one unit of resource. Plot (b) shows that the resource flows from items to users according to each item's degree and the resource on each user can be calculated by Eq. (2). Plot (c) shows the resource flows back to items based on each user's degree and the final resource on each item is calculated by Eq. (3).

Although the MD has proved its worth, there are still some problems. MD issues a recommendation with implicit feedback that contains only a binary value, such as 1 for a positive recommendation and 0 for a negative recommendation. Explicit feedback is neglected by the MD. In this paper, we will improve the mass diffusion model through a combined similarity strategy that integrates implicit feedback and explicit feedback. In addition, we will take into account some uncertainty about the similarity of user preferences that we will express using the Dempster-Shafer theory of belief functions.

3.1 Modeling similarities and uncertainty of preferences

Similarity measurement among users is a key part of user preference assessment. In the recommended systems, we always assume that the user will accept suggestions or options from other similar users. Therefore, how to measure the resemblance between users gets a lot of attention lately.

The similarity of cosine is a widely used approach in evaluating user preferences based on explicit feedback, e.g. between user *i* and *j*, the similarity of cosine is defined as:

$$Cos(i,j) = \frac{\sum_{l=1}^{n} R_{il} R_{jl}}{\sqrt{\sum_{l=1}^{n} R_{il}^2} \sqrt{\sum_{l=1}^{n} R_{jl}^2}} \tag{4}$$

where R_{il} and R_{jl} are rating points on item *l* rated by user *i* or *j*; *n* is the number of items that are evaluated by both users. The cosine similarity value is found in [0; 1], since the score rating is greater than 0. The similarity of the cosine measures the angle between two user evaluation vectors, where the larger value of the cosine similarity shows a closer relationship between the two users [1].

Assuming that nodes *i* and *j* are two users and node *l* is an item, it becomes part of the step 1 of the MD model. The resource allocation process can then be considered a one-step random walk on a bipartite network of user items starting with their common neighbors. Thus, step 1 in the MD model is equivalent to the process of measuring the similarity between two users when the initial resource on the items is one unit.

Further to express uncertainty about the similarity of user preferences, we define a frame of discernment Ω, which divides the similarity values into ten subintervals of the interval [0, 1].

In the context of social recommendation, the feeling of uncertainty of the user is often triggered by the diversity of the evaluation of his friends. This type of uncertainty can be modeled by Shannon entropy. Suppose user *u* has a final set of friends $f(u)$ and a discrete rating scale of *r* is enabled in the rating system. Let $c_j^u(i)(j \in [1,2,\dots,r]$ represent the number of ratings that are given to both the item and the friends at the rating scale *j*. The probability that a user expects *j* to be rated by his friends for *i* is given by:

$$p(u,i,j) = \frac{c_j^u(i)}{\sum_{j=1}^{r} c_j^u(i)} \tag{5}$$

According to Shannon entropy, the feeling of uncertainty about an item is due to the different ratings provided by u's friends:

$$S(u,i) = -\sum_{j=1}^{r} p(u,i,j)\log p(u,i,j) \tag{5}$$

Higher $S(u, i)$ indicates that u is more uncertain about i. In addition to the various ratings provided by friends of u, the interest in the item is also influenced by the total number of reviews provided by friends of u. Intuitively, when the rating varieties of two movies are the same, the user u will be more attracted to that one which is more popular among his friends. For example, a suggested movie that has aroused a dispute among a more friends will be more attractive than the one that has aroused a dispute only between two friends. Compared to the first case, this is a situation where the uncertainty regarding the allocation of the item's rating is high due to lack of evidence. Such uncertainty can be modeled by the Dempster-Shafer theory [2, 3, 10]. Accordingly, the distribution uncertainty of distribution is:

$$D(u,i) = \frac{r}{\sum_{j=1}^{r} c_j^u(i)+r} \tag{6}$$

A lower $D(i, i)$ value indicates a higher certainty about the distribution of ratings provided by the company's friends, and one step further shows that more evidence is available to support different views. Therefore, the feeling of uncertainty caused by the variety of ratings should be weighted by the inverse value of $D(u, i)$ given by:

$$N(u,i) = (1 - D(u,i)) \cdot S(u,i)$$

Therefore, $N(u, i)$ represents interest in the item and caused by a feeling of uncertainty as a result of information about the rating of friends.

To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer [24]. It is expressed by equations [102]:

$$^{\alpha}m(A) = \begin{cases} (1-\alpha)\times m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha)\times m(\Omega) & \text{if } A = \Omega \end{cases} \tag{7}$$

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

$$B(u,i) = Cos(i,j)(1 - D(u,i)) \cdot S(u,i) \tag{8}$$

The modified uncertain similarity diffusion model (USDM) is designed by integrating explicit and implicit feedback and uncertainty in preferences of users. In an MD, the resource is distributed based on the degree of each node, resulting in impersonal recommendations. While step 1 in the MD considers only implicit feedback, the USDM includes joint feedback and uncertainty calculation. Two-stage resource allocation process for the USDM model can be expressed in the following way:

Step 1: Assume that each item collected by the target user i is assigned to one unit of the initial resource. So the amount of resource will be distributed to user j with included uncertainty is defined as:

$$\phi_{ij} = \sum_{l=1}^{n} \frac{a_{il}a_{jl}B(u,i)}{\sum_{k=1}^{m} k_{kl}B(u,i)}\varphi_l \tag{9}$$

$\sum_{k=1}^{m} k_{kl}B(u,i)$ means the sum of the similarity between the target user i and all users who have collected the item, which is normalization. Feedback is used to calculate user uncertainty in the resource allocation process.

Step 2: The resource allocated to users will be returned to the items to complete the resource redistribution process. We intend to take into account both the user's degree and the degree to increase the variety of recommendations. The λ parameter controls the influence of the user level and the degree in this step. Assuming that an item receives a resource from users, the final resource on the item can be defined as:

$$\phi_{ir} = \sum_{j=1}^{m} \frac{a_{jr}}{k_r^{\lambda}k_j^{1-\lambda}}\phi_{ij} \tag{10}$$

We replace Eq. (9) to Eq. (10) to develop the ultimate model of our proposed method and project the relocation process of resources into the network of item items:

$$\phi_{ir} = \sum_{j=1}^{m} \frac{a_{jr}}{k_r^{\lambda}k_j^{1-\lambda}} \sum_{l=1}^{n} \frac{a_{il}a_{jl}B(u,i)}{\sum_{k=1}^{m} k_{kl}B(u,i)}\varphi_l \tag{11}$$

Finally, all items are sorted by their final source, and then they are for the target user *i*.

4 Results of our experiments

For our experiments, we used datasets called "Social Recommendation Data" from web site https://cseweb.ucsd. edu/~jmcauley/datasets.html#social_data [20]. These datasets include ratings as well as social (or trust) relationships between users. Data are from LibraryThing (a book review website) and epinions (general consumer reviews).

Dataset	Number of users	Number of items	Number of ratings/feed-back	Number of social rela-tions
Librarything	73,882	337,561	979,053	120,536
Epinions	41,554	112,991	181,394	181,304

Table 1 Basic statistics of datasets used in experiments

To evaluate our model, we used some widely used metrics to measure the accuracy of referral systems. In this paper we present one of them. It is about the accuracy of the recommendation. Precision (Pre) is an important metric for evaluating prediction in recommendation systems that measures the fraction of top-L recommended items that a target user consumes. Mathematically, the average Pre value for all users is defined as:

$$Pre = \frac{1}{m}\left(\sum_{i=1}^{m} \frac{D_i(L)}{L}\right) \tag{12}$$

where $D_i(L)$ is the number of recommended items consumed by the user *i* in the test file when the length of the recommendation list is *L*.

We compared the performance of our recommendation algorithm USDM with two recommendation methods used in e-commerce application:

1. PopRank is a basic recommendation algorithm with implicit feedback. It provides a recommendation list based on items' popularity. The most popular item will be at the top.

2. UserCF is a classic collaborative filtering method based on the cosine similarity between users. This method assumes the target user will accept the opinions from the most similar users.

Recommendation method	**Datasets**	
	Librarything	**Epinions**
PopRank	0.2584	0.3011
UserCF	0.3156	0.3865
USDM	0.3048	0.4236

Table 2 Recommendation performance of USDM and two chosen recommendation methods

We used two real datasets to evaluate our proposed method. The performance of our method is then compared with the two recommended algorithms used. Selected experimental results are presented in Table 2. The USDM algorithm we propose corresponds to the performance of the standard UserCF Recommendation. In further experiments we want to focus on the optimization of the λ parameter (Eq. 10), which has an additional impact on the accuracy of the recommendations. Also, we performed certain experiments within our subjects with students [4]. Here, too, the results indicated that the accuracy of our method is more than comparable to the other methods of recommendation used.

5 Conclusion

This paper proposes an uncertain similarity diffusion model (USDM) to improve the performance of recommendation. Experiments on two datasets demonstrated the good performance of our method.

In our future work, we will continue our research with experiments with our method in order to improve its performance, and with comparisons with other recommendation methods.

References

[1] Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178(1), 37–51.

[2] Beranek, L., Tlusty, P. & Remes, R. (2010). The Usage of Belief Functions for an Online Auction Reputation Model. In: M. Houda, J. Friebelova (Eds.) *Proceedings of the 28th Conference on Mathematical Method in Economy 2010* (pp. 49–54). Economical faculty. University of South Bohemia, Ceske Budejovice, Czech Republic.

[3] Beranek, L. (2011). The use of uncertain reasoning for the determination of the user's trustfulness on online auctions. In R. Matoušek (Ed.), *Proceedings of the 17th International Conference on Soft Computing: Mendel 2011* (pp. 360–365). Brno: Brno University of Technology.

[4] Beranek, L. (2015). The attitude of the college students to entrepreneurial skills development in the subject E-commerce. *Informatics in education*, 14(1), 1–12.

[5] Chen, L.-J., Zhang, Z.-K., Liu, J.-H., Gao, J. & Zhou, T. (2017). A vertex similarity index for better personalized recommendation. *Physica A: Statistical Mechanics and its Applications*, 466, 607–615.

[6] Gogna, A. & Majumdar, A. (2015). A comprehensive recommender system model: Improving accuracy for both warm and cold start users. *IEEE Access*, 3, 2803–2813.

[7] Liu, M., Pan, W., Liu, M., Chen, Y., Peng, X. & Ming, Z. (2017). Mixed similarity learning for recommendation with implicit feedback. *Knowledge-Based Systems*, 119, 178–185.

[8] Lu, J., Wu, D., Mao, M., Wang, W. & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, 74, 12–32.

[9] Lü, L. & Liu, W. (2011). Information filtering via preferential diffusion. *Physical Review E*, 83(6), 066119.

[10] Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

[11] Verma, C., Hart, M., Bhatkar, S., Parker-Wood, A. & Dey, S. (2016). Improving scalability of personalized recommendation systems for enterprise knowledge workers. *IEEE Access*, 4, 204–215.

[12] Vidmer, A. & Medo, M. (2016). The essential role of time in network-based recommendation. *EPL* (*Europhysics Letters*), 116(3), 30007.

[13] Wang, H., Wu, S., Pan, P., Zhang, L. & Chen, L. (2017). Towards large-scale social networks with online diffusion provenance detection. *Computer Networks*, 114, 154–166.

[14] Wang, W., Zhang, G. & Lu, J. (2016). Member contribution-based group recommender system. *Decision Support Systems*, 87, 80–93.

[15] Wu, D., Zhang, G. & Lu, J. (2015). A fuzzy preference tree-based recommender system for personalized business-to-business e-services. IEEE Transactions on Fuzzy Systems, 23(1), 29–43.

[16] Xiong, F. & Li, Z.-Y. (2017). Effective methods of restraining diffusion in terms of epidemic dynamics. Scientific Reports, 7, 6013.

[17] Xu, Z., Zhang, F., Wang, W., Liu, H. & Kong, X. (2016). Exploiting trust and usage context for cross-domain recommendation. IEEE Access, 4, 2398–2407.

[18] Yu, F., Zeng, A., Gillard, S. & Medo, M. (2016). Network-based recommendation algorithms: A review. *Physica A: Statistical Mechanics and its Applications*, 452, 192–208.

[19] Zhang, Y. C., Blattner, M. & Yu, Y. K. (2007). Heat conduction process on community networks as a recommendation model. *Physical Review Letters*, 99(15), 154301

[20] Zhao, T., McAuley, J. & King, I. (2015). Improving latent factor models via personalized feature projection for one-class recommendation. Conference on Information and Knowledge Management (CIKM), 2015, (pp. 821–830), Melbourne, Australia: ACM.

[21] Zhou, T., Ren, J., Medo, M. & Zhang, Y.-C. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, 76, 046115.

Review of several numerical approaches to sensitivity measurement of the Black–Scholes option prices

Jiří Hozman[1], Tomáš Tichý[2], Dana Černá[3], Aleš Kresta[4]

Abstract. The efficient and robust numerical pricing of options plays an important role in financial engineering. Since the option price value depends on several underlying parameters, it is desired to know, apart from its value, also how the option price is sensitive to the changes in the underlying parameters, such as asset prices, volatilities, interest rates and times to maturity. These sensitivity measures are named after Greek letters and simply called the Greeks of an option.
In this contribution, we provide a review and comparison of several approaches tested on the well-known benchmark of the Black–Scholes model for plain vanilla options. Specifically, we present three methodological concepts arising from finite differences, the discontinuous Galerkin approach and wavelet method.
The potential of each of the approaches is demonstrated within a simple empirical study with an emphasis on the number of basis functions. From the practical point of view we evaluate the options at several underlying nodes and compare these values to the reference and analytical ones. Simultaneously we investigate the convergence property and orders of the schemes presented.

Keywords: option pricing, Black–Scholes equation, finite differences, discontinuous Galerkin method, wavelet method

JEL Classification: C44
AMS Classification: 65M06, 65M60, 35Q91, 91G60

1 Introduction

The options and the financial models needed for their valuation have acquired increasing popularity during the last decades. The simplest options, in other words vanilla options, as well as a number of more complex options, are currently the most frequently used financial instruments. Therefore, valuing different types of option contracts together with their rigorous *sensitivity measurement* plays a very important role in modern financial theory and practice. This essential task is typically formulated by various partial differential equations (PDEs), for which analytical solutions can be derived in a closed form only under very restrictive conditions. Therefore, a wide range of studies have focused on the numerical realization of the option pricing problem, ranging from stochastic simulations [6] to numerical solutions of PDEs using finite difference methods [5] or variational techniques [13].

In this paper, we continue in our previous research on numerical option pricing via modern numerical techniques from [8], where *discontinuous Galerkin* (DG) method [12] and *wavelet* approach [2] are employed. However, instead of the price itself, here we study the sensitivities of option prices on selected factors, such as the underlying asset price, time to maturity, volatility, etc. These factors are crucial especially when sensitivities of large option portfolios are analyzed and risk limits are considered. In this paper we specifically focus on a relatively simple problem of an option sensitivities under Black and Scholes (BS) setting since in such case a closed-form solution is available. In particular, we analyze the computational error in the standard L^2-norm compared to its discrete version, l^2-norm, from the *finite difference* method [5] and study the behavior of the sensitivity functions.

The results obtained illustrate the potency of modern numerical methods in option pricing problems and serve as the cornerstone for their extended application to other option contracts and advanced option pricing models since the treatment is analogous to the BS framework. The emphasis is obviously placed on the valuation of options with more complex payoff functions (i.e., exotic options), e.g., Asian option contracts on two assets [9], and stochastic models for the standard parameters of option pricing models, e.g., one-factor stochastic volatility models [10].

[1] Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz
[2] Department of Finance, VSB-TU Ostrava, Sokolska 33, 701 21, Ostrava, Czech Republic, tomas.tichy@vsb.cz
[3] Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, dana.cerna@tul.cz
[4] Department of Finance, VSB-TU Ostrava, Sokolska 33, 701 21, Ostrava, Czech Republic, ales.kresta@vsb.cz

2 Option Pricing and Sensitivity Measurement

Options are nonlinear types of financial derivatives, which give the holder the right (but not the obligation) to buy the underlying asset in the future (at maturity time) at the prespecified exercise price. Simultaneously, the writer of the option has to deliver the underlying asset if the holder asks. Therefore, the valuation is quite challenging.

The standard market model proposed independently by Black and Scholes [1] and Merton [11] is valid, in its basic form, only under idealized market conditions, including perfect market with the underlying asset price following log-normal distribution without any dividends and its returns having constant volatility σ. Following these arguments the price of European vanilla option V written on a single asset S with maturity T is the unique solution of the following backward partial differential equation with the terminal condition

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \quad \text{for } S \geq 0, T \geq t \geq 0 \quad \text{with } V(S,T) = \begin{cases} \max(S - \mathcal{K}, 0), & \text{for a call,} \\ \max(\mathcal{K} - S, 0), & \text{for a put,} \end{cases} \tag{1}$$

and has the asymptotic behavior

$$\lim_{S\to 0+} V(S,t) = \begin{cases} 0, \\ \mathcal{K}e^{-r(T-t)}, \end{cases} \qquad \lim_{S\to+\infty} V(S,t) = \begin{cases} S - \mathcal{K}e^{-r(T-t)}, & \text{(call)} \\ 0. & \text{(put)} \end{cases} \tag{2}$$

The symbol r stands for the riskless interest rate and $\mathcal{K}$ denotes the strike price. The problem (1)–(2) has analytical solutions given by the BS formula as

$$V(S,t;\sigma,r,\mathcal{K},T) = \begin{cases} S\Phi(d_1) - \mathcal{K}e^{-r(T-t)}\Phi(d_2), & \text{for a call,} \\ \mathcal{K}e^{-r(T-t)}\Phi(-d_2) - S\Phi(-d_1), & \text{for a put,} \end{cases} \tag{3}$$

where $d_1 = [\ln(S/\mathcal{K}) + (r + \sigma^2/2)(T-t)]/[\sigma\sqrt{T-t}]$, $d_2 = d_1 - \sigma\sqrt{T-t}$ and Φ stands for the cumulative distribution function of the standard normal distribution.

Recall that the option value (3) depends on several underlying parameters and it is obvious that any parameter change should consequently influence it. The sensitivity analysis and measurement show how significant these changes will be – since Greek letters are commonly used to denote such sensitivity measures, we often call them *the Greeks* of an option. The availability of analytical solutions in the closed form, such as those for European call and put options as presented in (3), implies the ability of deriving corresponding closed form representations for the sensitivity measures as well, see [14].

In the rest of the paper we focus only on the first-order Greeks that are represented by the first derivatives with respect to the underlying parameters, namely

$$\begin{aligned} \Delta_V &= \frac{\partial V}{\partial S} = \eta\Phi(\eta d_1), & \Theta_V &= \frac{\partial V}{\partial t} = -\eta r\mathcal{K}e^{-r(T-t)}\Phi(\eta d_2) - \frac{\sigma S\phi(d_1)}{2\sqrt{T-t}}, \\ \mathcal{V}_V &= \frac{\partial V}{\partial \sigma} = S\sqrt{T-t}\,\phi(d_1), & \rho_V &= \frac{\partial V}{\partial r} = \eta(T-t)\mathcal{K}e^{-r(T-t)}\Phi(\eta d_2), \end{aligned} \tag{4}$$

where ϕ is a density function of the standard normal distribution and η indicates the option type ($\eta = 1$ for call and $\eta = -1$ for put option, respectively).

The simplest sensitivity measure is Delta (Δ_V) that measures the sensitivity of the theoretical option value with respect to the changes in the underlying asset's price. The rate of change of an option value due to the passage of time is measured by Theta (Θ_V), also referred to as an option's time decay. Further, in volatile markets the value of some option positions can be particularly sensitive to changes in the volatility of the underlying asset price returns, and thus Vega ($\mathcal{V}_V$), the derivative of the option value with respect to such volatility, should be taken into account. The last of the first-order Greeks considered here is Rho (ρ_V) that measures the sensitivity to the riskless interest rate. For further research we refer the interested reader to the book [14].

3 Numerical Valuation of Options and Calculation of Greeks

In this section, we introduce the *discrete problem* to option valuation within the framework of the common finite difference methods and two relatively novel numerical techniques based on the DG method and wavelet approach, which both fall into the class of variational methods. These new approaches represent a very powerful tool for the numerical simulation of option valuation, since they allow us better capturing of some features of different

options under various market conditions with respect to the discretization of the computational domain as well as the order of the polynomial approximation. The presented pricing methodologies related to numerical solving of the BS equation (1) require truncation of the spatial domain to a bounded interval $\Omega = (0, S_{\max})$, where $S_{\max} > 0$ stands for the sufficiently large asset price. To impose the option values at both endpoints of the domain Ω the relation (2) is employed. The derivation of each numerical technique is presented very briefly due to the lack of space and thus we refer to related papers and books cited below for the detailed description.

Finite Difference Method

The finite difference (FD) method, as one of the simplest approximations of partial differential equations (see [5]), replaces partial derivatives of (1) in Ω by suitable (finite) differences, namely

$$\frac{\partial V(S,t)}{\partial S} \approx \frac{V(S+h,t) - V(S-h,t)}{2h}, \qquad \frac{\partial^2 V(S,t)}{\partial S^2} \approx \frac{V(S+h,t) - 2V(S,t) + V(S-h,t)}{h^2}, \tag{5}$$

where h is a mesh step of the spatial grid given by the partition nodes $0 = S_0 < S_1 \ldots < S_N = S_{\max}$. Further, we can proceed in a similar way for the case of the temporal derivative $\partial V(S,t)/\partial t$. Let $0 = t_0 < t_1 < \ldots < t_M = T$ be a partition of the interval $[0, T]$ with the constant time step $\tau = T/M$, then taking into account backward time running, we have the approaximation $\frac{\partial V(S,t)}{\partial t} \approx -\frac{V(S,t-\tau)-V(S,t)}{\tau}$.

Now, after substituting the selecting suitable finite differences for particular partial derivatives in pricing equation (1), we also replace $V(S_k, t_m)$ by $V_{k,m}$, which states for approximation of the option value at price S_k and time t_m, then:

$$\frac{V_{k,m} - V_{k,m-1}}{\tau} + rkh\frac{V_{k+1,m} - V_{k-1,m}}{2h} + \sigma^2 k^2 h^2 \frac{V_{k+1,m} - 2V_{k,m} + V_{k-1,m}}{2h^2} = rV_{k,m}, \tag{6}$$

with $k = 1, \ldots, N-1$ and $m = 0, \ldots, M$. The values for $k = 0$ and $k = N$ are given by (2). In this way we can depict complete structure of option prices for all time states, including the initial states $V_{k,M}$ given by payoff and the final states $V_{k,0}$, and selected underlying asset prices S_k. Subsequently, we can rearrange terms in (6) as follows:

$$V_{k,m-1} = a_k V_{k-1,m} + b_k V_{k,m} + c_k V_{k+1,m}, \tag{7}$$

where $a_k = -\tau(rk - \sigma^2 k^2)/2$, $b_k = 1 - \sigma^2 k^2 \tau - r\tau$ and $c_k = \tau(rk + \sigma^2 k^2)/2$. In other words, the relation (7) can be read as that the option value at the time t_{m-1} and the central state S_k is equal to the weighted average (specified by coefficients a_k, b_k, and c_k) of option values at the previous time level t_m for three states of the underlying asset price (S_{k-1}, S_k, S_{k+1}).

Discontinuous Galerkin Method

Assume the same partitions of the space-time domain $\Omega \times (0, T)$ as for the finite difference method. Let $U_h^m \in S_h^p$ be the approximation of the solution $V(\cdot, t_m)$ and the set S_h^p be the finite-dimensional space of piecewise polynomial functions of order p, constructed over the partition of Ω with mesh size h, for a complete overview see [12]. The discrete solution within the DG framework is computed by the following scheme that reads: Find $U_h^m \in S_h^p$, $m = 0, \ldots, M-1$, such that the following conditions are satisfied:

$$\left(U_h^m, v_h\right) + \tau \mathcal{A}_h(U_h^m, v_h) = \left(U_h^{m+1}, v_h\right) - \tau \ell_h(v_h)(t_m) \quad \forall\, v_h \in S_h^p, \tag{8}$$

$$(U_h^M, v_h) = (V(\cdot, T), v_h) \quad \forall\, v_h \in S_h^p, \tag{9}$$

where $(\cdot, \cdot)$ denotes the inner product in $L^2(\Omega)$, the bilinear form $\mathcal{A}_h(\cdot, \cdot)$ stands for the DG semi-discrete variant of the degenerate parabolic partial differential operator from (1), accompanied with penalties and stabilizations, and the form $\ell_h(\cdot)(t)$ balances Dirichlet boundary conditions from (2), for more details see [7].

The scheme (8)–(9) is well known as the *implicit Euler scheme*, which is practically unconditionally stable and gives the first order convergence in time. Further, note that the equation (8) results into a sequence of linear algebraic problems. The existence and uniqueness of the discrete solution are guaranteed under the ellipticity of the form $(\cdot, \cdot) + \tau \mathcal{A}_h(\cdot, \cdot)$ on the left-hand side of (8), cf. [9].

The discrete problem (8) is equivalent to a system of linear algebraic equations at each time level and can always be expressed in a matrix form, see [9]. Let $B = \{\varphi_k\}_{k=1}^{\mathrm{DOF}}$ denote the basis of the space S_h^p with DOF degrees of freedom, then the discrete solution at each time level $t_m \in [0, T]$ can be written as a linear combination of basis functions, more precisely as

$$U_h^m(S) = \begin{cases} \sum_{k=1}^{\mathrm{DOF}} \xi_k^m \varphi_k(S), & \text{for } S \neq S_l, \\ \frac{1}{2}\sum_{k=1}^{\mathrm{DOF}} \xi_k^m (\varphi_k(S^+) + \varphi_k(S^-)), & \text{for } S = S_l, \end{cases} \quad 1 \le l \le N-1, \tag{10}$$

where symbols S^+ and S^- stand for the one-sided limits. The discrete solution U_h^m is identified with the coefficient vector $\mathbf{u}^m = \{\xi_k^m\}_{k=1}^{\mathrm{DOF}} \in \mathbb{R}^{\mathrm{DOF}}$ with respect to the basis B. Then (8) reads

$$(\mathbf{M} + \tau\mathbf{A})\,\mathbf{u}^m = \mathbf{M}\mathbf{u}^{m+1} - \tau\mathbf{f}^m, \tag{11}$$

where the terminal vector $\mathbf{u}^M$ is given by U_h^M arising from (9). The system matrix in (11) is a composition of the mass matrix $\mathbf{M}$ and the matrix $\mathbf{A}$ arising from the bilinear form $\mathcal{A}_h$ and the right-hand side of (11) contains the vector $\mathbf{f}^m$ which enforces the fulfillment of boundary conditions at time level t_m, i.e., $\mathbf{M} = \{(\varphi_j, \varphi_i)\}_{i,j=1}^{\mathrm{DOF}}$, $\mathbf{A} = \{\mathcal{A}_h(\varphi_j, \varphi_i)\}_{i,j=1}^{\mathrm{DOF}}$ and $\mathbf{f}^m = \{\ell_h(\varphi_i)(t_m)\}_{i=1}^{\mathrm{DOF}}$. Since the system matrix is non-symmetric the restarted GMRES solver is incorporated into the numerical procedure.

Wavelet Method

First, we briefly introduce the concept of a *wavelet basis*. We assume that H is a suitable space from a weak formulation of option pricing problem (see [7]) and that $\mathcal{J}$ is an index set such that each index $\lambda \in \mathcal{J}$ has the form $\lambda = (j, k)$, where $|\lambda| = j$ denotes the level. Set $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\}$ is called a wavelet basis for space H, if Ψ is a Riesz basis of H and has a hierarchical structure consisting of scaling functions and wavelets; the supports of ψ_λ are bounded with factor $2^{-|\lambda|}$ and wavelet basis functions have vanishing moments. For more details see, e.g., [3], where the linear and quadratic spline wavelet bases are employed.

The introduced notation is analogous to that in the previous section; specifically, $W^m(S)$ denotes $V(S, t_m)$. We again use the implicit Euler scheme to discretize equation (1) in time; and for discretization with respect to the spatial variable S, we use the wavelet method. The *adaptive wavelet method* differs from classical approaches, because it is based not on local error estimates but on thresholding wavelet coefficients.

Let Ψ be a wavelet basis of the appropriate space H, then we can expand the weak solution $W^m(S)$ in the wavelet basis as $W^m = \sum_{\lambda\in\mathcal{J}} \omega_\lambda^m \psi_\lambda$ and substitute it into a variational formulation. Accordingly, we obtain the bi-infinite system $\mathbf{B}\mathbf{w}^m = \mathbf{g}^{m+1}$ with the solution $\mathbf{w}^m = \{\omega_\lambda^m\}_{\lambda\in\mathcal{J}}$, matrix $\mathbf{B} = \{B_{\mu,\lambda}\}_{\mu,\lambda\in\mathcal{J}}$ and the right-hand side $\mathbf{g}^{m+1} = \{g_\mu\}_{\mu\in\mathcal{J}}$, defined as

$$B_{\mu,\lambda} = (\psi_\lambda, \psi_\mu) + \tau\mathcal{B}(\psi_\lambda, \psi_\mu), \quad g_\mu^{m+1} = (W^{m+1}, \psi_\mu) - \tau(q^m, \psi_\mu), \quad \mu, \lambda \in \mathcal{J}, \tag{12}$$

where $\mathcal{B}(\cdot,\cdot)$ stands for the bilinear form representing the spatial differential operator in a weak formulation and q^m balances the boundary conditions at time t_m, see [7]. We solve the resulting system in the following steps:

1. Choose the time step τ and the number of basis functions DOF.
2. Compute the vector of coefficients $\mathbf{w}^M$ for the payoff function and $\mathbf{w}_h^M = \mathrm{COARSE}(\mathbf{w}^M, \mathrm{DOF})$.
3. For $m = M-1, \ldots, 2, 1, 0$, compute the right-hand side $\mathbf{g}^{m+1}$ and calculate $\mathbf{w}^m = \mathrm{GMRES}(\mathbf{B}, \mathbf{g}^{m+1}, \mathbf{w}_h^{m+1})$ and $\mathbf{w}_h^m = \mathrm{COARSE}(\mathbf{w}^m, \mathrm{DOF})$.

In this algorithm, $\mathbf{w}_h^m = \mathrm{GMRES}(\mathbf{B}, \mathbf{g}^{m+1}, \mathbf{w}_h^{m+1})$ means that $\mathbf{w}_h^m$ is the solution of the system with matrix $\mathbf{B}$ and the right-hand side $\mathbf{g}^{m+1}$ using the GMRES with initial vector $\mathbf{w}_h^{m+1}$. The routine $\mathbf{w}_h^m = \mathrm{COARSE}(\mathbf{w}^m, \mathrm{DOF})$ consists of thresholding, that is, taking DOF elements of vector $\mathbf{w}^m$, which are the highest in absolute value, and setting the others to zero. Each iteration of the GMRES requires the multiplication of infinite-dimensional matrix $\mathbf{B}$ with the finite-dimensional vector. We compute this operation approximately following the method in [4].

Approximate Evaluation of Greeks

In order to illustrate the robustness of the presented approach, the numerical scheme (11) is used not only for the evaluation of option prices, but also their sensitivity measures. Considering the polynomial approximation at least of the first order (linear), Delta can be directly computed from the derivatives of the basis functions $\{\varphi_k'\}_{k=1}^{\mathrm{DOF}}$ using the relation (10) as

$$\Delta_V(t_m) \approx \frac{\partial U_h^m}{\partial S} = \sum_{k=1}^{\mathrm{DOF}} \xi_k^m \varphi_k'(S). \tag{13}$$

On the other hand, the remaining Greeks are numerically computed using the central finite difference, i.e.,

$$\Theta_V(t_m) \approx \frac{U_h^{m+1} - U_h^{m-1}}{2\tau}, \quad \mathcal{V}_V(t_m, \sigma) \approx \frac{U_h^m(\sigma+\delta) - U_h^m(\sigma-\delta)}{2\delta}, \quad \rho_V(t_m, r) \approx \frac{U_h^m(r+\delta) - U_h^m(r-\delta)}{2\delta}, \tag{14}$$

where $0 < \delta \ll 1$ and τ are sufficiently small values. This approach involves solving the option valuation problem twice to obtain the solution in two different stages $t \pm \tau$, $\sigma \pm \delta$ and $r \pm \delta$, respectively. Note that this approach provides only the pointwise approximation with respect to the underlying parameters t, σ, and r. For the wavelet method we proceed similarly as above. In the case of the FD approach the relation (13) is naturally replaced by the central finite differences.

4 Comparative Experiment

The experimental analysis provided in this section shows extensive comparison of finite diference, discontinuous Galerkin and wavelet methods in connection with the plain vanilla option price sensitivity measurement of the first order within the BS setting. In particular, we consider data in line with [8], where the valuation problem of vanilla put option at the same market data (DAX options) has been analyzed. We consider here the case of intermediate maturity of 193 days close to ATM options with current underlying value being $S_{\text{ref}} = 4\,715.879$ and strike price $\mathcal{K} = 4\,700$. The fixed parameters of the model are the riskless interest rate $r = 0.039$ and the volatility $\sigma = 0.4422$, which is derived from true option prices observed at the market (i.e., implied volatility approach).

Numerical approximation is crucially related to the discretization of the computational domain Ω, its length is deliberately chosen as eight times the strike price to suppress the influence of the inaccurate Dirichlet boundary condition. Together with this, we choose the time step $\tau = 1/3600$ so that the effect of time discretization on numerical results is negligible.

For a more detailed comparison, DG and wavelet methods are considered in the form of linear as well as quadratic approximation. The quality of the approximation can be easily observed by comparing the numerical results with the theoretical prices according to BS formula. Therefore, at $t = 0$, we compute relative error e_{L^2} measured in L^2-norm over the whole computational domain (for DG and wavelet methods) and its discrete version e_{l^2} (for FD method), i.e.,

$$e_{L^2} = \frac{\|Z_h^0(S) - V(S,0)\|_{L^2}}{\|V(S,0)\|_{L^2}}, \quad e_{l^2} = \frac{\|Z_h^0(S) - V(S,0)\|_{l^2}}{\|V(S,0)\|_{l^2}}, \tag{15}$$

where Z_h^0 denotes the approximate solution obtained by one of three numerical approaches and $V(S,t)$ is the analytical solution given by the BS formula (3), respectively. The formulas (15) can be subsequently extended to the calculation of the relative errors e_{L^2} and e_{l^2} for selected sensitivity measures using (4) and (13)–(14).

According to the theoretical results from [12] and [2], it is known that the DG and wavelet techniques produce (in general) optimal convergence of spatial derivatives of approximate solutions, all measured in L^2-norm for sufficiently regular problems. Therefore, the similar convergence results could be expected for the remaining first order derivatives.

The calculations are performed on a sequence of the consecutive uniformly refined meshes with linear as well as quadratic basis functions. The particular sensitivity measures are computed by (13) and using the central finite differences (14) with $\tau = 1/3600$ and $\delta = 0.002$, respectively. Corresponding relative errors are apparent from Figure 1. It is easy to see that relative errors in L^2-norm decrease with the mesh refinement, i.e., proportionally to DOF. More precisely, the order of convergence for the first-order Greeks corresponds asymptotically to the order of a polynomial approximation. One can easily conclude that all methods considered are quite comparable with respect to relative errors and we can not clearly choose the best method from them. As concerns the quadratic basis functions, the relative errors exhibit the similar behavior as for the linear basis functions. More precisely, in line with the theory, the convergence of Delta values in L^2-norm is optimal, i.e., quadratic for the approximation with quadratic basis functions. As expected, the results for the remaining first-order Greeks are of the same quality and their orders of accuracy are consistent with the Delta ones. However, the smallest L^2-errors are in most cases produced by wavelet method.

5 Conclusion

This paper presents selected convergence results of three approaches to numerical sensitivity measurement of European vanilla options under Gaussian setting, namely discontinuous Galerkin, wavelet and finite difference methods. For the comparison analysis, mid term ATM on German stock market index have been selected. The results do not show any significant differences among particular techniques and measures. However, the wavelet method provides the best approximation properties. Regarding the DG method, its main advantages lie in the possibility of an easy usage of complex payoff functions or discrete sampling in more advanced pricing models, which will be objective of further research.

Acknowledgements

Supported by SP2019/5, an SGS research project of VSB-TU Ostrava, and the ESF in "Science without borders" project, reg. nr. CZ.02.2.69/0.0/0.0/16_027/0008463 within the Operational Programme Research, Development and Education.

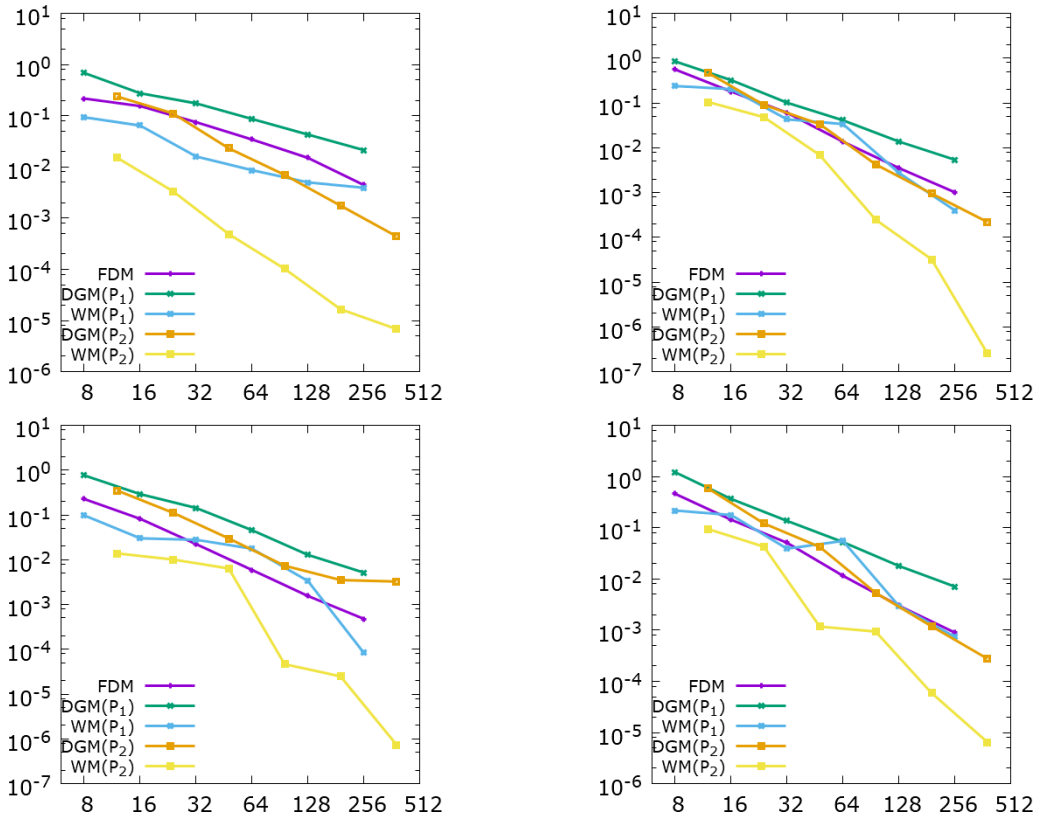


Figure 1 Comparison of relative errors of the Delta (*top left*), Theta (*top right*), Vega (*bottom left*), and Rho (*bottom right*) values for particular methods. The horizontal axis represents the degrees of freedom.

References

[1] Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–659.

[2] Cohen, A. & Dahmen, W., & Devore, R. (2002). Adaptive Wavelet Methods II — Beyond the Elliptic Case. *Foundations of Computational Mathematics*, 2, 203–245.

[3] Černá, D. & Finěk, V. (2011). Construction of optimally conditioned cubic spline wavelets on the interval. *Advances in Computational Mathematics*, 34, 219–252.

[4] Černá, D. & Finěk, V. (2013). Approximate multiplication in adaptive wavelet methods. *Central European Journal of Mathematics*, 11, 972–983.

[5] Duffy, D.J. (2006). *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach*. Chichester: John Wiley & Sons.

[6] Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. New York: Springer.

[7] Hozman, J., et al. (2018). *Robust Numerical Schemes for Pricing of Selected Options*. SAEI, Vol. 57., Ostrava: VŠB-TU Ostrava.

[8] Hozman, J. & Černá, D. & Holčapek, M. & Tichý, T. & Valášek, R. (2018). Review of modern numerical methods for a simple vanilla option pricing problem. *ER-CEREI*, 21, 21–30.

[9] Hozman, J. & Tichý, T. (2017). DG method for numerical pricing of multi-asset Asian options — the case of options with floating strike. *Applications of Mathematics*, 62, 171–195.

[10] Hozman, J. & Tichý, T. (2018). DG framework for pricing European options under one-factor stochastic volatility models. *Journal of Computational and Applied Mathematics*, 344, 585–600.

[11] Merton, R.C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4, 141–183.

[12] Rivière, B. (2008). *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. SIAM, Philadelphia.

[13] Topper, J. (2005). *Financial Engineering with Finite Elements*. Chichester: John Wiley & Sons.

[14] Ugur, Ö. (2008). *An Introduction to Computational Finance*. London: Imperial College Press.

Impact of tax changes on economic growth of Slovakia

Stanislav Kováč[1]

Abstract. The power of government in changing the direction and the speed of convergence to the steady state of country's economic growth is indisputable. This paper deals with the tax changes in Slovakia and their impact on economic growth. We present commonly used methods to compute the structural parameters which appear in analysis based on a linear approximation around the steady state of Cass-Koopman model. We also present a simple algorithm which can be later reproduce for similar studies. In empirical part we display a graphical representation of convergence simulation with emphases on the limits of tax rates. Our study can serve as a monitor of current state and as a suggestion of government possibilities in tax area.

Keywords: taxation, linear approximation, convergence

JEL Classification: C50, E62
AMS Classification: 91B62

1 Introduction

Since the 1980s, analysis of economic growth (captured in GDP per capita) has become one of the most active fields in economic research. During this relatively small period, academicians have been able to come up with many different factors of growth, for example:

- fiscal and monetary policy (Easterly and Rebelo [4]);
- spatial factors (Chocholatá [3]);
- entrepreneurial activity (Stel, Carree and Thurik [11]).

In this paper we focus on direct and indirect taxes and their impact on economic growth in Slovakia using modified Cass-Koopman model (following Novales, Fernández and Ruíz [10]) . There are dozens of papers with similar goal for different countries using various methods, beginning with a negative correlation between fiscal policy and economic growth to more complex models. Lee and Gordon [8] analyzed the effects of corporate and personal tax rates on economic growth of 70 countries using panel regression. They found out that cut in the corporate tax rate by 10 % would raise the annual growth rate by one to two percentage points. Tosun and Abizadeh [13] showed the stronger response of economic growth on direct than indirect taxes for OECD countries. With comparable conclusions came several authors, e.g., Bâzgan [1].

The paper has the following structure: in the next paragraph we present Cass-Koopman model and its linear approximation around steady states, in the third one we make an empirical analysis based on presented methodological concept and the last paragraph concludes.

2 Methodology

The main assumption of Cass-Koopman model is the usage of a single agent which represents private sector including consumers and firms. The representative agent is interested in maximizing her/his welfare which is described by a constant risk aversion utility function. We also assume agent accessibility to a production technology using labor and capital to produce a single good.

The representative agent solves the optimization problem in form

$$\max_{\{c_t,k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}-1}{1-\sigma} \tag{1}$$

subject to

$$(1+\tau_t^c)c_t + (1+n)k_{t+1} - (1-\delta)k_t = \left(1-\tau_t^y\right)f(k_t), \tag{2}$$

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, stanislav.kovac@euba.sk.

where β stands for a discount factor $\beta \in (0,1)$; c, k are consumption and capital per capita respectively; σ comes with the risk aversion utility function as a coefficient of aversion $\sigma \in (0, \infty)$; τ_t^c, τ_t^y represents the time-dependent tax rates on consumption and production; n, δ are constant rates of population growth and capital depreciation, finally $f(\cdot)$ is a production function per capita of the Cobb-Douglas type $f(k_t) = Ak_t^{\alpha}, \ \alpha \in (0,1)$.

The Lagrangian is the summation of current and discounted future values of utility and **discounted** budget constraint (2)

$$\mathcal{L}(\{c_t, k_{t+1}, \lambda_t\}) = \sum_{t=0}^{\infty} \beta^t \left[\frac{c_t^{1-\sigma} - 1}{1-\sigma} + \lambda_t \begin{bmatrix} \left(1 - \tau_t^y\right) A k_t^{\alpha} - (1 + \tau_t^c) c_t \\ -(1+n) k_{t+1} + (1-\delta) k_t \end{bmatrix} \right]. \tag{3}$$

Necessary conditions with respect to c_t, k_{t+1} and the transversality condition have the form

$$\beta^t [c_t^{-\sigma} - \lambda_t (1 + \tau_t^c)] = 0, \tag{4}$$

$$-\beta^t \lambda_t (1+n) + \beta^{t+1} \lambda_{t+1} \left[\left(1 - \tau_{t+1}^y\right) A \alpha k_{t+1}^{\alpha - 1} + (1 - \delta) \right] = 0, \tag{5}$$

$$\lim_{t \to \infty} \beta^t \lambda_t k_{t+1} = 0. \tag{6}$$

Equation (4) can be rewritten

$$\lambda_t = \frac{c_t^{-\sigma}}{(1 + \tau_t^c)} \tag{7}$$

and after the substation of (7) into (5) we get Keynes-Ramsey condition

$$c_{t+1} = \left[\frac{\beta}{(1+n)} \frac{(1 + \tau_t^c)}{(1 + \tau_{t+1}^c)} \left[\left(1 - \tau_{t+1}^y\right) A \alpha k_{t+1}^{\alpha - 1} + (1 - \delta) \right] \right]^{\frac{1}{\sigma}} c_t. \tag{8}$$

Solving the system of nonlinear equations (2), (8) and (6) characterizes the optimal trajectories of capital and consumption, which can be solved by numerical methods. With k_0 and c_0 given we can compute k_1 with (2) and c_1 with (8), however it will almost certainly not converge. Therefore, we have to set the stability conditions where we assume constant consumption (c_{ss}), capital (k_{ss}) and production (y_{ss}) over time as well as constant tax rates τ^y, τ^c. Taking these assumptions into account, we obtain the steady state of capital from relation (8)

$$k_{ss} = \left[\frac{(1 - \tau^y) A \alpha}{\frac{1+n}{\beta} - (1 - \delta)} \right]^{\frac{1}{1-\alpha}} \tag{9}$$

and after applying assumptions with (9) into (2) we obtain the steady state of consumption

$$c_{ss} = \frac{1}{1 + \tau^c} [(1 - \tau^y) A k_{ss}^{\alpha} - (n + \delta) k_{ss}]. \tag{10}$$

We use a linear approximation around the steady state which comes from the Taylor expansion up to linear term. Then, the budget constraint (2) with the Cobb-Douglas production function and constant tax rates has the linear approximation in form

$$k_{t+1} - k_{ss} = \frac{1}{\beta} (k_t - k_{ss}) - \frac{1 + \tau^c}{1+n} (c_t - c_{ss}). \tag{11}$$

Similarly, we use linear approximation to Keynes-Ramsey condition (8) with constant tax rates

$$(c_{t+1} - c_{ss}) = \frac{1}{\sigma} \Omega_{ss}^{\frac{1}{\sigma} - 1} \left[\frac{\beta}{(1+n)} (1 - \tau^y) A \alpha (\alpha - 1) k_{ss}^{\alpha - 2} \right] c_{ss} (k_{t+1} - k_{ss}) + \Omega_{ss}^{\frac{1}{\sigma}} (c_t - c_{ss}) \tag{12}$$

where

$$\Omega_{ss} = \frac{\beta}{(1+n)} [(1 - \tau^y) A \alpha k_{ss}^{\alpha - 1} + (1 - \delta)]. \tag{13}$$

Because of the steady state ($c_{t+1} = c_t$) it comes from Keynes-Ramsey condition (8) that $\Omega_{ss} = 1$. Then we get

$$(c_{t+1} - c_{ss}) - \frac{1}{\sigma}\frac{\beta}{(1+n)}(1-\tau^y)A\alpha(\alpha-1)k_{ss}^{\alpha-2}c_{ss}(k_{t+1} - k_{ss}) = (c_t - c_{ss}). \quad (14)$$

These linear approximations (11) and (14) in vector matrix form

$$\begin{pmatrix} 1 & 0 \\ M & 1 \end{pmatrix}\begin{pmatrix} k_{t+1} - k_{ss} \\ c_{t+1} - c_{ss} \end{pmatrix} = \begin{pmatrix} \frac{1}{\beta} & -\frac{1+\tau^c}{1+n} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} k_t - k_{ss} \\ c_t - c_{ss} \end{pmatrix}, \quad (15)$$

where

$$M = -\frac{1}{\sigma}\frac{\beta}{(1+n)}(1-\tau^y)A\alpha(\alpha-1)k_{ss}^{\alpha-2}c_{ss}. \quad (16)$$

The formula (15) can be written as

$$\begin{aligned} \boldsymbol{B_0 z_{t+1}} &= \boldsymbol{B_1 z_t} \\ \boldsymbol{z_{t+1}} &= \boldsymbol{D z_t}, \end{aligned} \quad (17)$$

where $\boldsymbol{D}$ is a transit matrix which spectral decomposition has the form $\boldsymbol{D = \Gamma \Lambda \Gamma^{-1}}$ where $\boldsymbol{\Gamma}$ is a matrix of eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. If we use $\boldsymbol{D = \Gamma \Lambda^t \Gamma^{-1}}$ instead, we can compute the whole trajectory with initial values of consumption and capital. We have

$$\begin{pmatrix} k_t - k_{ss} \\ c_t - c_{ss} \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}\begin{pmatrix} \lambda_1^t & 0 \\ 0 & \lambda_2^t \end{pmatrix}\begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}\begin{pmatrix} k_0 - k_{ss} \\ c_0 - c_{ss} \end{pmatrix}, \quad (18)$$

where

$$\lambda_{1,2} = \left((d_{11} + d_{22}) \pm \sqrt{(d_{11} + d_{22})^2 - 4(d_{11}d_{22} - d_{12}d_{21})}\right)/2, \quad (19)$$

$$\boldsymbol{x}^T = (1, (\lambda_1 - d_{11})/d_{12}), \quad (20)$$

$$\boldsymbol{y}^T = (1, (\lambda_2 - d_{11})/d_{12}), \quad (21)$$

$$\boldsymbol{u}^T = ((\lambda_2 - d_{11})/(\lambda_2 - \lambda_1), -(\lambda_1 - d_{11})/(\lambda_2 - \lambda_1)), \quad (22)$$

$$\boldsymbol{v}^T = (-d_{12}/(\lambda_2 - \lambda_1); d_{12}/(\lambda_2 - \lambda_1)). \quad (23)$$

The eigenvalues of transit matrix determine the rate of growth. Following a saddle point structure, the eigenvalues satisfy $|\lambda_1| > 1/\sqrt{\beta}$ and $|\lambda_2| < 1$. It is obvious that only the second root is stable. The transversality condition holds only if a coefficient related to the unstable eigenvalue in (18) is zero. If $x_1 = 1$, then

$$u_1(k_0 - k_{ss}) + v_1(c_0 - c_{ss}) = 0 \quad (24)$$

or with substitution (22) and (23) we get

$$c_0 - c_{ss} = \frac{\lambda_2 - d_{11}}{d_{12}}(k_0 - k_{ss}). \quad (25)$$

In similar manner, keeping the stable eigenvalue in (18) we have

$$k_t - k_{ss} = y_1\lambda_2^t[u_2(k_0 - k_{ss}) + v_2(c_0 - c_{ss})], \quad (26)$$

$$c_t - c_{ss} = y_2\lambda_2^t[u_2(k_0 - k_{ss}) + v_2(c_0 - c_{ss})], \quad (27)$$

which can be simplified into

$$c_t - c_{ss} = \frac{\lambda_2 - d_{11}}{d_{12}}(k_t - k_{ss}). \quad (28)$$

The algorithm of computation can be summarized into 5 steps:

1. setting the values (estimation) of parameters: $\beta, \delta, A, \alpha, n, \sigma, \tau^y, \tau^c$ and initial value of capital per capita k_0;
2. computing the steady state levels of capital (9) and consumption (10);
3. computing the transition matrix (17) and its eigenvalues (19), eigenvectors (20), (21);
4. numerical evaluation of initial value of consumption c_0 according to (25);
5. obtaining k_1 from (2) and c_1 from (28), repeating for all *t*.

3 Empirical results

In the following section we will examine the effects of direct and indirect tax changes on economic growth in Slovakia. We proceed in the manner of presented algorithm.

The first step requires to estimate some structural parameters. We use yearly data of variables retrieved from Eurostat, macroeconomic database of Slovak National Bank and Gallup World Poll from 2000 to 2017. All parameters, used methods and the initial value of capital per capita are presented in Table 1.

Parameters	Slovakia	Method/Source
β	0.963113	$1/(1+E[r_t])$, r_t- real interest rate
δ	0.057426	$E[KC_t/K_t]$, KC_t - capital consumption, K_t - capital
A	26.590000	$A = E[y_t/k_t^{\alpha}]$
σ	0.920000	Gandelman and Hernandez-Murillo [5]
α	0.580814	$E[AWS_t], AWS_t$ - adjusted wage share
n	0.000450	$E[(N_{t+1}-N_t)/N_t], N_t$ - population
τ^y	0.210000	accace.com/ [6]
τ^c	0.200000	accace.com/ [6]
$k_{0=2017}$	37 403.890000	K_{2017}/N_{2017}

Table 1 Structural parameters and initial value of capital per capita

We assume the prices of production factors to be equal to their marginal products therefore a wage share on production is equal to $(1-\alpha)$ and capital share is equal to α. Parameter α can be quantified through national accounts. It is not correct to include all incomes into the consideration except from capital income, specifically in the case of self-employed persons. We use an adjusted wage share (aws_t) with formula

$$AWS_t = (W_t/Y_t)(L_t^1/L_t^0) \tag{29}$$

where W_t is a remuneration of employees, Y_t is GDP, L_t^0 stands for a number of employees and L_t^1 represents a number of workers which is equal to the summation of the number of employees and the number of self-employed persons. The wage share is the average of adjusted wage share which in case of Slovakia from 2000 to 2017 is 0.419186, hence the capital share on production is 0.580814.

The only unspecified parameter from the Cobb-Douglas is a technological parameter A. We modify our production function where all known parameters are put on the left side and unknown on the right one

$$y_t/k_t^{\alpha} = A \tag{30}$$

where variables are divided with the number of population and the constant can be calculated as the average of the division.

The problem arises with a capital stock series such as does not exist for Slovakia, nor does a detailed break down of gross fixed capital investment into sectors in constant prices. The method employed is a simplified perpetual inventory model (PIM) using the value of the capital stock supplied by the Statistics Office of the Slovak Republic as the initialization reference which according to Livermore [9] is 4 525.02 bn SKK in the end of 1998 in the current prices. After the conversion with an investment deflator (88.67 in 1998) and SKK-EUR exchange

rate we get the capital stock 169.713 bn EUR which we will refer to the estimate of 1999 as well. Then, capital stock for next periods are calculated with PIM

$$K_{t+1} = K_t + I_{t+1} - KC_{t+1} \tag{31}$$

where K_t is a capital stock in t, I_{t+1} is a gross fixed capital formation in $t+1$ and KC_{t+1} represents a capital consumption in $t+1$. The average ratio of capital consumption on capital is a constant rate of capital depreciation (0.057426), although it has an increasing trend we abstract from this trend. And using formula (30) we get technological constant $A = 26.59$. We also get the initial value of capital per capita which in our analysis refers to the year 2017 (37 403.89 €). It is important to mention that according to Szomolanyi, Lukáčik and Lukáčiková [12] there is an evidence of leaving Cobb-Douglas type of production function in case of Slovakia.

In estimating risk aversion, we use the estimation presented by Gandelman and Hernandez-Murillo [5] who used Gallup World Poll. They did not estimate the parameter for Slovakia itself, however, it is not the problem because as a benchmark we use the average coefficient among developed countries which has the value 0.92.

For a direct income tax (tax rate on production), we use a corporate income tax which is levied at a rate of 21 %. We do not consider special taxes on corporate income which fall under special legislation. In case of the indirect tax (tax rate on consumption), we use the value added tax (VAT) at a standard rate 20 %. The choice for VAT as a tax rate on consumption comes from the assumption that agent represents households as well as companies. Revenue of VAT was approximately 43 % of government budget revenue in 2017 and at the same time revenue of corporate income tax was 18 % of the budget. Using these two we captured more than a half of Slovak budget revenue.

In the next step we calculate the steady states of consumption and capital per capita. The steady state of capital per capita has the value 104 096.54 € according to formula (9) and consumption per person 9 346.27 € using formula (10). Considering Cobb-Douglas production function we also can compute the steady state of production per capita which has the value 21 823.06 € (in 2017 Slovak GDP per capita is 15 010.35 €).

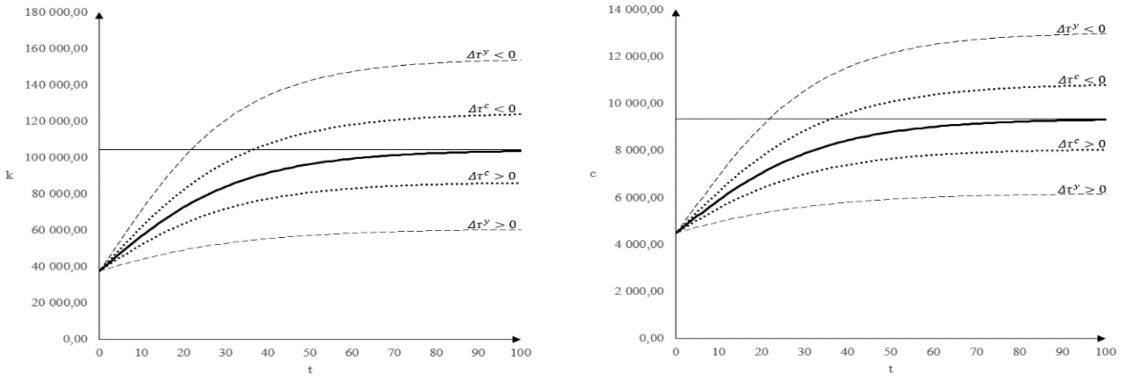


Figure 1 Trajectories of capital per capita (left) and consumption per capita (right) in Slovakia

In Figure 1 we can see the convergence trajectories of consumption per capita and capital per capita under current state in Slovakia. The solid line represents all variables holding constants. The higher impact has a tax rate on income compare to a tax rate on consumption in both cases (consumption and capital). If we decrease the tax rate on income about 50 %, the economy convergences to the initial steady state approximately about 1.75-times faster than it would be done through the same relative change of tax rate on consumption. Interestingly, by the time consumption per capita with the half of income tax rate has reached the initial steady state, the cumulative government budget revenue has the value of 70 % of unchanged tax rate revenue, in other words, 50 % tax-rate-cut led to 30 % revenue drop. This proves Tosun and Abizadeh [13] finding. However, they vastly differ from Lee and Gordon [8]. They claim the average effect of 10 % cut in corporate tax leads to the raise of economic growth about 1 percentage point. We find this effect to hold for 50 % cut in corporate income tax.

Because we have abstract from time dependency, it is important at this point to consider plausible changes of steady states related to the trends of structural parameters. As most of the developed countries, Slovakia faces to declining population curve as well, which leads to upward movement of steady states of variables per capita. The opposite and stronger effect has increasing capital depreciation which move downward steady states.

4 Conclusion

In this paper we derived a linear approximation around steady state of neoclassical Cass-Koopman model where we also presented a simple algorithm which can be applied to similar studies. We applied a methodological concept on measuring the impact of taxation on economic growth of Slovakia. There were also shown the possible outcomes of tax rate changes and structural parameters on steady states. We showed that government can accelerate the convergence trough the tax on corporate income with revenue drop, however, the rational attitude of government would be done through the increasing VAT which has a smaller effect on downward movement of steady states and still led to higher government revenue.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic - Vega grant No. 1/0294/18 "Short-run and long-run dynamic analysis of economic development of European post-communist countries and regions" and Vega grant No. 1/0248/17 "Analysis of Regional Disparities in the EU based on Spatial Econometric Approaches".

References

[1] Bâzgan, R. M. (2018) The impact of direct and indirect taxes on economic growth: An empirical analysis related to Romania. In A. M. Dima (ed), *Proceedings of the International Conference on Business Excellence 2018* (pp. 114-127). Warsaw: Sciendo.

[2] Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. In *Review of Economic Studies*, 32 (3), 233-240.

[3] Chocholatá, M. (2017). Relationship of the regional growth and volatility: spatial econometric analysis for the V4 regions. In R. Nemec & L. Chytilová (eds), *Proceedings of the 12th International Conference on Strategic Management and its Support by Information Systems 2017* (pp. 239-246). Ostrava: VŠB - Technical University of Ostrava.

[4] Easterly, W. & Rebelo, S. (1993). Fiscal policy and economic growth. In *Journal of Monetary Economics*, 32 (3), 417-458.

[5] Gandelman, N. & Hernandez-Murillo, R. (2015). Risk aversion at the country level. In *Federal Reserve Bank of St. Louis Review*, 97 (1), 53-66.

[6] https://accace.com/ [cited 2019-04-01].

[7] Koopmans, T. C. (1965). On the concept of optimal economic growth. In *Study Week on The Econometric Approach to Development Planning* (pp. 225-287). Amsterdam: North-Holland Publishing.

[8] Lee, Y. & Gordon, R. H. (2005). Tax structure and economic growth. In *Journal of Public Economics*, 89 (5-6), 1027-1043.

[9] Livermore, S. (2004*). An Econometric Model of the Slovak Republic.* [Online]. Available at: https://www.finance.gov.sk/en/finance/institute-financial-policy/working-papers/the-econometric-model-slovak-republic/ [cited 2019-04-01].

[10] Novales, A. & Fernández, E. & Ruíz, J. (2009). *Economic Growth. Theory and Numerical Solution Methods*. Berlin: Springer-Verlag.

[11] Stel, A. & Carree, M. & Thurik, R. (2005). The effect of entrepreneurial activity on national economic growth. In *Small Business Economics*, 24 (3), 311-321.

[12] Szomolányi, K. & Lukáčik, M. & Lukáčiková, A. (2017). Long-run elasticity of the substitution in the Slovak economy. In M. Baćović et al. (eds), *ETRENOVA Conference Proceedings (September 7, 2017)* (pp. 228-232). Zagreb: Udruga za promicenja inovacija i istraživanja u ekonomiji "IRENET".

[13] Tosun, M. S. & Abizadeh, S. (2005). Economic growth and tax components: an analysis of tax changes in OECD. In *Applied Economics*, 37 (19), 2251-2263.

Interest Rate Modelling: Maximum Likelihood Estimation of One-Factor Short-Rate Models

Tomáš Rusý[1], Kamil Kladívko[2]

Abstract. The maximum likelihood method is known to be efficient at estimating fully parametric models. One-factor short-rate models belong to this class, but surprisingly the maximum likelihood method is not extensively used for estimating them. We believe it is a consequence of the current method's failure to determine the value of the short rate without justifying the calculation procedure, which often leads to a poor fit of the observed curve, making it difficult to interpret. In this paper, we propose a way to consider all observed yields at one time and extract the value of the short rate jointly from the entire yield curve. This could be done thanks to a general description of the construction of the likelihood function of a time series of observed yields. The method identifies the models under the real-world measure and hence it is suited not only for pricing, but also for prediction of interest rates. We illustrate the use of such an approach on the popular Hull - White model.

Keywords: interest rate, short rate models, maximum likelihood.

JEL classification: C53
AMS classification: 91G30, 62P05

1 Introduction

Anticipating future evolution of interest rates is of the utmost importance in the world of finance. Financial institutions exchange vast amount of money. Therefore a good model showing accurate predictions of future yields is priceless in assessing their investment opportunities. In the following paper, we will present a way how one can obtain reliable estimates of parameters within affine-term-structure one-factor short-rate models, which are often used for that purpose. These models use rather a theoretical concept of unobserved short rate, which implies the distribution of future yields. The most frequent approach of calibrating such models is matching current prices of interest rate derivatives. However, such methods do not allow to obtain estimates of the market price of risk as it is already included in the prices of derivatives. Another drawback arises in emerging markets when we might not observe enough traded derivatives so the obtained parameter values might be quite unstable and not trustworthy.

An alternative way can be seen in the maximum likelihood estimation suggested by Chen and Scott [3]. The idea of such a method is to utilise the implied distribution of the short rate given by the model and to construct the likelihood function from the density of the distribution. However, as the short rate is not observed on the market, it remains unclear what data should be used in the likelihood function. Chen and Scott [3] derived the value of the short rate from a model implied relationship between the short rate and a single yield-to-maturity. They had to specify in advance the corresponding tenor they used and therefore they eliminated the effect of other yields on the short rate. In contrast, Cerny [2] and Fergusson and Platen [5] used strictly three month and one year yield as a proxy, which appears to us rather inept. None of these approaches use information observed in the entire yield curve. We will generalise the construction of the likelihood function from Chen and Scott [3] so that it allows the utilisation of such information. Moreover, the construction will be very flexible at modifying model assumptions to fit various datasets, while being applicable to a general class of one-factor models.

We begin in Section 2 with theoretical foundations, where the generalised construction of the likelihood will be presented. Section 3 is devoted to the introduction of the Hull - White model of Hull and White [7], which will be subject to analysis in later section. There, we also describe parameter settings, which is applied in the numerical example. Its results are then presented in Section 3.

[1]Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75, Praha 8, rusy@karlin.mff.cuni.cz
[2]Örebro University School of Business, kamil.kladivko@oru.se

2 Construction of the Log-Likelihood Function

In the following section, we will present the idea of the construction of the likelihood function of one-factor short-rate models. We will introduce the "pricing" errors and specify, how they enter the relationship between the model implied yield curve and the observed rates. Subsequently, we will show how one can derive the density of the yield curve for such parametrisations. From there, the likelihood function can be formulated.

2.1 One-Factor Short-Rate Models

First, we would like to shortly mention the difference between the real world measure and the risk neutral measure, which are generally used for specifying the short rate dynamics. The latter includes the market price of risk — a quantity derived by Vasicek [8], which summarizes how much investors charge for holding risky positions. Their relationship was exploited in number of places, e.g. in a book of Brigo and Mercurio [1]. In our aim to express the density of historical rates, we have to work with the model specified under the less frequently used real world measure. In general, we consider the short rate to have dynamics

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t),$$

where $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are some sufficiently smooth deterministic functions and $W(t)$ is a Brownian motion under the real world measure. The interpretation of such an equation is that the change of the short rate is driven by one deterministic factor and one random factor. We denote the model implied conditional (at time t given time $s < t$) density of the short rate as $f_r(r_t|r_s)$. Numerous short-rate models have already been described and analysed. We should mention the Vasicek [8] or the Hull - White model [7], for which f_r is shown to be a density of a normal distribution. On the other hand, the Cox - Ingersoll - Ross [4] model implies that r_t given r_s has a non-central χ^2 distribution. We restrict ourselves to the models which imply so called affine term structure of zero-coupon bond prices. Formally, it means that the time t model implied value of a zero-coupon bond paying 1 unit at time $t + \tau$ can be written as

$$P(t, t+\tau) = \exp\{A(t, t+\tau) - B(t, t+\tau)r_t\}$$

for some arbitrary functions $A(t, t+\tau)$ and $B(t, t+\tau)$. Functions A and B can also depend on the short rate model parameters, but for simplicity these will be omitted in the notation. Annualized and continuously compounded time t yield-to-maturity $t + \tau$, which we denote $z_t(\tau)$ can be expressed as

$$z_t(\tau) = -\frac{1}{\tau}\log P(t, t+\tau) = a(t, \tau) + b(t, \tau)r_t, \tag{1}$$

where $a(t, \tau) = -A(t, t+\tau)/\tau$ and $b(t, \tau) = B(t, t+\tau)/\tau$. The model definition implies that the short rate r_t is given information at time s a random variable. Hence conditioned on the knowledge of information at time s, $z_t(\tau)$ is also a random variable, as $z_t(\tau)$ is a function of r_t. Let us have n yields $z_t(\tau_1), \ldots, z_t(\tau_n)$ with times-to-maturity $\tau_1, \ldots, \tau_n$ and let us denote the following quantities:

$$\begin{aligned} z_t &= \big(z_t(\tau_1), \ldots, z_t(\tau_n)\big)^T, \\ A_t &= \big(a(t, \tau_1), \ldots, a(t, \tau_n)\big)^T, \\ B_t &= \big(b(t, \tau_1), \ldots, b(t, \tau_n)\big)^T. \end{aligned}$$

Using this, obtain vectorized equation (1) as

$$z_t = A_t + B_t r_t. \tag{2}$$

Equation (2) summarizes the model relationship between the short rate and the considered yields to maturity. It is uniquely determined by the model dynamics, its parameters, times-to-maturity $\tau_1, \ldots, \tau_n$ and by time t. It could be interpreted as the model implied yield curve. In the Vasicek model, vectors A_t and B_t are not time dependent on the other hand for example in the Hull - White model A_t is time dependent.

2.2 Pricing Errors

From (1) it follows, that given a single yield-to-maturity τ, we can reconstruct the value of the short rate only by inverting the relationship. However, if we do it for another yield, we would obtain different value of the short rate

for the same time t. This theoretical relationship in reality does not hold exactly. For that reason, we introduce a concept that the yields are measured with pricing errors. These will summarize fluctuations of real yields around the model implied yield curve. So let ε_t be a random vector which represents this error. It is moreover assumed to be independent of the short rate process. Next, we will need to redistribute these $n-1$ errors between our n observations of the yield curve. Chen and Scott [3] assumed, that one observation (the one with the shortest maturity) was observed without any error, while every other observation was associated with its own error. We generalise such a formulation. Let us define C as an $n \times n-1$ matrix which distributes these $n-1$ errors into n components of yields. From (2), we get

$$y_t = A_t + B_t r_t + C\varepsilon_t = z_t + C\varepsilon_t = A_t + \begin{pmatrix} B_t & C \end{pmatrix} \begin{pmatrix} r_t \\ \varepsilon_t \end{pmatrix}. \tag{3}$$

Let us compare this equation to (2). We can see, that we have added to the model-derived relationship as stated in (1) the pricing errors $C\varepsilon_t$ to make the relationship realistic. As we said, z_t can be interpreted as a theoretical (model based) yields-to-maturity, while y_t represents the observed yields to maturity. Our aim will be to find a suitable model for the random variable y_t. Matrix $\begin{pmatrix} B_t & C \end{pmatrix}$ from (3) will be denoted T_t. It is an $n \times n$ matrix which defines the transformation between the observed yields and the vector of the short rate and errors.

2.3 The Density and the Log-Likelihood

We are particularly interested in the conditional density of y_t at time t given time s. We will make use of the relationship in (3) and use a theorem which describes relationship of densities of transformed random variables. We have $y_t = g(r_t, \varepsilon_t)$ where g is a linear transformation, so it is monotone and also the determinant of the transformation is given by the determinant of matrix T_t. Let us denote the conditional density of y_t given y_s as $f_y(y_t|y_s)$ and the density of errors as $f_\varepsilon(\varepsilon_t)$. Then, with the use of the independence of the short rate and errors, we can express

$$f_y(y_t|y_s) = \frac{1}{|\det T_t|} f_r(\hat{r}_t|r_s) f_\varepsilon(\hat{\varepsilon}_t), \tag{4}$$

where

$$\begin{pmatrix} \hat{r}_t \\ \hat{\varepsilon}_t \end{pmatrix} = T_t^{-1}(y_t - A_t). \tag{5}$$

Therefore, in order to be able to express the density of the observations of the yield curve, we need to specify the matrix of the transformation T_t, which must be regular. Moreover we need to know its inverse and its determinant. The values of A_t, B_t which appear in (5) along with the density f_r can be derived from the short rate dynamics. The density of errors f_ε is also specified by ourselves as one of the assumptions of the model. In general, we will be looking for such a density so that one can assume that the fitted errors $\hat{\varepsilon}_t$ were generated from it. Finally, we will need to specify the structure of the matrix C.

As we aim to use the maximum likelihood method to obtain the estimates of the parameters of the model, we will need to derive the log-likelihood function. Let us assume that our observations are measured at m periods, at times $t_1 < \cdots < t_m$. In other words, we observe n-dimensional time series $\{y_{t_j}\}_{j=1}^m$. Given a vector of parameters $\theta \in \Theta$, the likelihood of the data can be expressed as

$$\begin{aligned} L(\theta; y_{t_1}, \ldots, y_{t_m}) &= f_{y_{t_1},\ldots,y_{t_m}}(y_{t_1}, \ldots, y_{t_m}), \\ &= f_y(y_{t_m}|y_{t_{m-1}}) \cdots f_y(y_{t_2}|y_{t_1}) f_y(y_{t_1}), \end{aligned}$$

where θ is the vector of the parameters of the model, consisting of parameters of the one-factor short-rate model and of the parameters of the error distribution. In the following, we will work only with the conditional likelihood function (conditioned on y_{t_1}). That can be written as

$$L(\theta; y_{t_1}, \ldots, y_{t_m}|y_{t_1}) = \prod_{j=1}^{m-1} f_y(y_{t_{j+1}}|y_{t_j}) = \prod_{j=1}^{m-1} \frac{1}{|\det T_{t_{j+1}}|} f_r(\hat{r}_{t_{j+1}}|r_{t_j}) f_\varepsilon(\hat{\varepsilon}_{t_{j+1}}).$$

From the above equation we can derive the form of the conditional log-likelihood function.

$$\ell(\theta) = -\sum_{j=1}^{m-1} \log|\det T_{t_{j+1}}| + \sum_{j=1}^{m-1} \log f_r(\hat{r}_{t_{j+1}}|r_{t_j}) + \sum_{j=1}^{m-1} \log f_\varepsilon(\hat{\varepsilon}_{t_{j+1}}), \tag{6}$$

where $\hat{r}_{t_j}$ and $\hat{\varepsilon}_{t_j}$ are calculated for each time t_j as in (5). Remember, that the vector of parameters θ, includes the parameters of the short rate model (these usually enter the formulas for A_t, B_t and f_r) and on the parameters of the error distribution (these enter formulas for f_ε). In some cases, the part of the likelihood corresponding to the errors can be maximised analytically so we obtain a profile log-likelihood function for the short-rate model parameters. This significantly helps in the numerical routines, which need to be used to obtain the estimates.

3 Empirical Data Analysis

We will apply the above-described method to the Hull - White model or sometimes called the Extended Vasicek model, which was developed by Hull and White [7] in such a way, that it fits perfectly to the observed market yield curve. Such property makes it popular between analysts for pricing interest rate derivatives.

3.1 The Hull - White Model

The Hull - White model is usually defined under the risk neutral measure $\mathbb{Q}$ with the following dynamics:

$$dr_t = \Big(\bar{\theta}(s,t) - \alpha r_t\Big)dt + \sigma dW^Q(t),$$

where $W^Q(t)$ is the standard $\mathbb{Q}$-Brownian motion and $s \leq t$ defines the time when we observe the yield curve. Parameter α stands for mean reversion factor and σ summarizes the volatility of the short rate r_t. Finally, $\bar{\theta}(s,t)$ is set such that the observed market prices at time s are fitted perfectly, i.e.

$$\bar{\theta}(s,t) = \frac{\partial f^M(s,u)}{\partial u}\Big|_{u=t} + \alpha f^M(s,t) + \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha(t-s)}\right), \qquad f^M(s,t) = -\frac{\partial \log P^M(s,u)}{\partial u}\Big|_{u=t},$$

where $f^M(s,t)$ is the market instantaneous forward rate from time s at time t. $P^M(s,t)$ is the time s market observed price of a zero coupon bond paying one at time t. See Brigo and Mercurio [1] for more details. For maximum likelihood estimation, real world measure $\mathbb{P}$ is required. Parametrisation under such a measure can be obtained via the Girsanov's theorem (see for example Harrison and Pliska [6]):

$$W^P(t) = W^Q(t) - \int_0^t \lambda du,$$

where W^P is a Brownian motion under the real world measure. There for simplicity, we assume that the market price of risk is constant. By solving the stochastic differential equation, we obtain that conditionally on time s, the short rate is under such a measure again normally distributed with mean and variance as

$$\mu_{t|s} = f^M(s,t) + \frac{\sigma^2}{2\alpha^2}\left(1 - e^{-\alpha(t-s)}\right)^2 - \frac{\sigma\lambda}{\alpha}\left(1 - e^{-\alpha(t-s)}\right), \qquad \sigma^2_{t|s} = \frac{\sigma^2}{2\alpha}\Big(1 - e^{-2\alpha(t-s)}\Big). \tag{7}$$

The functions $a(t,\tau)$ and $b(t,\tau)$ defined in (1), which specify the relationship between the short rate and the yield with time-to-maturity τ will be denoted by $a^{HW}_{t|s}(\tau)$ and $b^{HW}(\tau)$ respectively.

$$b^{HW}(\tau) = \frac{1 - e^{-\alpha\tau}}{\alpha\tau}, \;\; a^{HW}_{t|s}(\tau) = F^M(s,t,t+\tau) - b^{HW}(\tau) f^M(s,t) + \tau\frac{\sigma^2 b^{HW}(\tau)^2}{4\alpha}\left(1 - e^{-2\alpha(t-s)}\right), \tag{8}$$

where function $F^M(s,t,T)$ represents the market observed forward rate at time s from time t with maturity T. Note that $b^{HW}(\tau)$ does not depend on the time index t.

3.2 Error Distribution

Finally, we need to specify the error distribution and the C matrix. We assume that the error distribution is centred normal with a regular variance-covariance matrix $\Sigma = \Sigma_{(n-1)\times(n-1)}$. For the matrix C, we propose to use one such that to the first $n-1$ yields, an own error is assigned, and to the last yield, sum of all errors divided by a factor $\sqrt{n-1}$ is assigned. This would imply the matrix T and also the inverse T^{-1} to be

$$T = \begin{pmatrix} b_1 & 1 & & 0 \\ \vdots & & \ddots & \\ b_{n-1} & 0 & & 1 \\ b_n & \frac{1}{\sqrt{n-1}} & \cdots & \frac{1}{\sqrt{n-1}} \end{pmatrix}, \quad T^{-1} = \frac{1}{a} \begin{pmatrix} -\frac{1}{\sqrt{n-1}} & \cdots & -\frac{1}{\sqrt{n-1}} & 1 \\ a + \frac{b_1}{\sqrt{n-1}} & & \frac{b_1}{\sqrt{n-1}} & -b_1 \\ & \ddots & & \\ \frac{b_{n-1}}{\sqrt{n-1}} & & a + \frac{b_{n-1}}{\sqrt{n-1}} & -b_{n-1} \end{pmatrix}, \tag{9}$$

where $b_1, \ldots, b_n$ are the elements of B as in (3) and $a = b_n - \sum_{i=1}^{n-1} b_i/\sqrt{n-1}$. The reason for such a choice is that the matrix is sparse, which allows us to calculate the inverse of the transformation, while it ensures suitable properties for the errors of observed yields. Mainly, it creates the last yield error as a symmetric combination of the $n-1$ errors and hence the properties of $C\varepsilon_t$ are similar to ε_t. We also have that $|\det T| = |a|$ and a necessary condition that $a \neq 0$. This does not appear to be a limiting restriction.

3.3 Results

In this section, we will present the results of the above-described model fitted to the PRIBOR dataset, which includes quoted ON (overnight), 2 week, $1, 2, 3, 6, 9$ month and one year rates along with $2, 3, \ldots, 10$ year rates which were bootstrapped from at-the-money swap rates. The series consists from monthly observations starting at 31st May 1999 and ending at 30th November 2016. We have selected 8 tenors for the analysis because of computational efficiency and needed detailed diagnostic of the model. These were 3 and 6 months, $1, 2, 3, 4, 5, 7$ and 10 years. To optimise the log-likelihood function, we used software R and the Nelder - Mead optimisation method. We faced no numerical problems when optimising the log-likelihood function. The maximum likelihood estimates of the short rate model parameters were the following:

$$\hat{\alpha} = 0.036963242, \qquad \hat{\sigma} = 0.005958489, \qquad \hat{\lambda} = 0.901977764. \tag{10}$$

Next, we investigated if we meet model assumptions. Their violation would in general result in an invalid statistical inference. In that case, we would face a question if we can believe in the model and in general how much the model

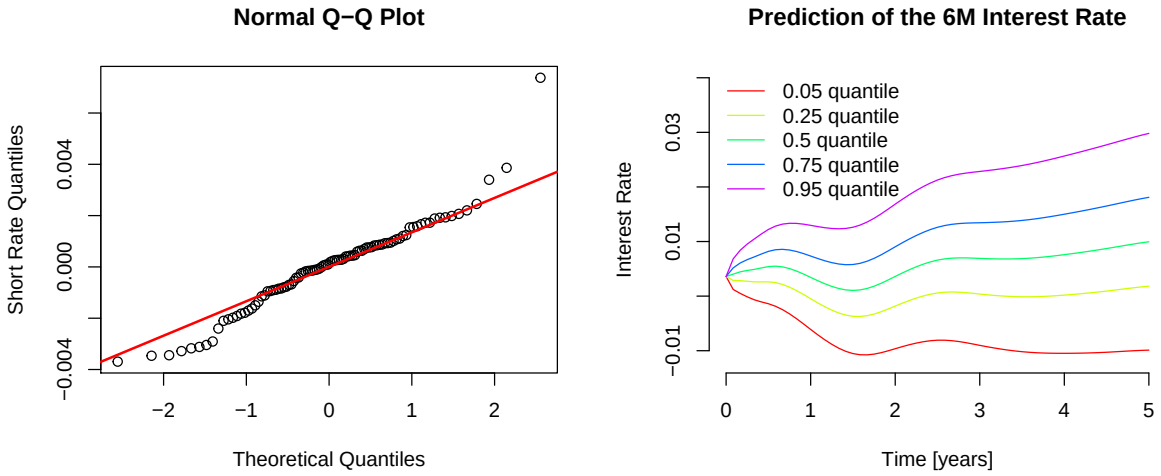


Figure 1: Prediction of the six-month interest rate as implied by the calibrated Hull - White model for the next 5 years and QQ plot of short rate movements.

parameters reflect their true meaning. The first testable assumption is whether the short rate follows a normal distribution. To asses the normality, one can have a look at the left figure in Figure 1, where a normal QQ plot of the short rate movements is shown. Shapiro-Wilk test for normality gives p-value of 0.0013. Hence we reject the hypothesis at any usually considered level of significance. The reason lies in the tails of the distribution, when changes in the short rate are sometimes "too large." This can be down to some artificial interventions, which have much greater effect than normal distribution can model. We also studied the autocorrelation of differences of the short rate and the *acf* and *pacf* plots did not suggest any significant autocorrelation in the series. For the errors, these have been assumed to be time independent and normal. For the normality, we applied Shapiro-Wilk test, which rejected marginal normality in two of the seven tests at 5% level of significance (for the third and the fifth error component). When we looked in more detail into the two time series, we found out that this is caused by an outlier in the fourth observation. This could suggest that the economy still did not completely overcome the changes in the crisis year 2008. When we left out the first four observations, marginal normality was not rejected by any of the tests. This could tell us that the departures from normality are not great. The *acf* and *pacf* plots supported our surmise and did not suggest any violation of the no time-correlation between errors.

4 Conclusion

The implications of short-rate models are immense and once a suitable parameter values are obtained, then, for example, the Hull - White model under the real world measure specifies the distribution of yield curve at any given time. This can be used for pricing, Value-at-Risk calculation or scenario generation. However, the maximum likelihood method has been barely used for estimating them, which is surprising given that it is known to be efficient in fully parametric models. We attribute this to the fact that it is complicated to extract the short rate as a latent factor and suggested possibilities were not performing well. For example Cerny [2] found that yields implied by a fitted Vasicek model were not close to the observed ones. This drawback basically ruled out any application, as it creates a model which is too far away from reality.

Our approach towards deriving the short rate can be seen as a breakthrough in the sense that it utilise information in the entire yield curve. This increases greatly the interpretation of the latent short rate process. Another advantage is that it can be applied in most of the markets, including the ones with illiquid market of interest rate derivatives. Finally, the fact that it identifies the model under both real world and risk neutral measures highlights the usefulness of the method and enlarges the number of situations where it can be applied.

Acknowledgements

The work was supported by the grants GAUK No. 258318 and GAČR No. 18-05631S.

References

[1] D. Brigo and F. Mercurio. *Interest Rate Models – Theory and Practice with Smile, Inflation and Credit (2nd ed. 2006)*. Springer Verlag, Berlin, 2001. ISBN 978-3-540-22149-4.

[2] J. Cerny. Calibration of interest rates. *WDS12 Proceedings of Contributed Papers: Part I – Mathematics and Computer Sciences*, 2013. Available at SSRN: https://ssrn.com/abstract=2302509.

[3] R.R. Chen and L. Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income*, 3:14–31, 1993.

[4] J.C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53: 385–407, 1985.

[5] K. Fergusson and E. Platen. Application of maximum likelihood estimation to stochastic short rate models. *Annals of Financial Economics*, 10:27–53, 2015.

[6] J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260, 1981.

[7] J. Hull and A. White. Pricing interest–rate derivative securities. *The Review of Financial Studies*, 3:573–592, 1990.

[8] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

Czech Economy under Foreign Exchange Intervention Regimes

Tomáš Oravec [1], Osvald Vašíček [2]

Abstract. After the economic recession with continued monetary policy easing, many European central banks have been forced to implement unconventional monetary policy instruments. In the case of the Czech economy, depreciation of domestic currency on foreign exchange market has been employed while maintaining nominal exchange rate against euro no stronger than the announced level. In the submitted paper, we inspect the small open Czech economy under several intervention regimes within DSGE model with occasionally binding constraints (OBC). In particular, we examine commitment, managed floating and real exchange rate rule intervention regimes in the economy constrained by zero lower bound at the same time. We provide economic behaviour for individual foreign exchange interventions based on impulse responses framework. We present responses to positive as well as negative unanticipated shocks hitting the economy.

Keywords: unconventional monetary policy, intervention, occasionally binding constraints, commitment, managed floating, real exchange rate rule.

JEL Classification: E37, E50
AMS Classification: 91B51, 91B64

1 Introduction

The period of the Great Recession has been very dramatic from the perspective of many developed countries. In order to mitigate economic consequences of the crisis, vast majority of central banks kept lowering their policy interest rates. This regular monetary policy option has been no longer feasible at the moment when nominal interest rates have been dropped at technically zero level. Since ordinary policy options have been exhausted, central banks around the globe have been adopting extraordinary actions. In the recent history, mostly quantitative easing and interventions on the foreign exchange (FX) market as unconventional monetary programmes have been employed. Because Czech economy is a small one, the latter option by means of domestic currency artificial depreciation has been introduced while maintaining nominal exchange rate with respect to Euro above the threshold of 27 CZK/EUR. Similar asymmetric commitments have been introduced by central bank in e.g. Switzerland (against Euro) or Israel (against Dollar).

The main research question is to assess impacts of such unconventional monetary policy both quantitatively and qualitatively. However, there have been papers estimating impacts of this monetary policy in the case of Czech Republic, let us mention papers Brůha & Tonner [2] or Lízal & Schwarz [8]. The approach used in this paper includes modern non-trivial modelling of constrained economy by means of occasionally binding constraints (OBC) introduced in Holden & Paetz [6] and technically corroborated in Holden [5]. Our work extends research presented in Fodor & Vašíček [4], Malovaná [9], Súkupová & Vašíček [12], Súkupová & Vašíček [13] and Tvrz *et.al.* [14] by incorporating and comparison of several intervention regimes into zero-level interest rates economic environment. We have estimated baseline (unconstrained) model on Czech economy data and simulated economic behaviour within respective intervention regimes afterwards.

The rest of the paper is structured as follows. Section 2 sketches model's structure. Section 3 provides mathematical methodology of OBC simulations. Empirical results are included in Section 4. Finally, Section 5 concludes.

2 Model

This section introduces basic features of Czech small open economy model. Selected model is mainly adapted from Alpanda *et.al.* [1], while microeconomic details follow Justiniano & Preston [7] and Steinbach *et.al.* [11]. Hence, Czech economy is a part of two-country DSGE model, where European Monetary Union (EMU) countries are

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, tomas.oravec@mail.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz

considered to be the foreign large economy. In general, agents' behaviour in present affects future environments. Some uncertainty arises, because specific processes in the economy are exposed to exogenous (random) shocks.

In the model, monopolistic competition is assumed, i.e. prices are set by private economic agents maximizing their objectives. In particular, there is a continuum of representative agents in households and firms. Firms consist of domestic and retail ones producing and selling their products. Households provide labour services for which they receive wages. In addition, households are assumed to smooth their own consumption over time, and hence can spend their income on bonds either.

Key ingredient of New-Keynesian DSGE models is nominal rigidities featured in compliance with Calvo [3]. That means, some of the agents face constraints on the frequency with which they can adjust prices of goods and services they provide. A consequence of nominal rigidities is short run non-neutrality of monetary policy, leading to variations in real interest rate. Nevertheless, all prices revert back to their natural equilibria in the long run.

Standard New-Keynesian framework is extended by risk-averse and short-sighed FX dealers and FX market according to Montoro & Ortiz [10]. Solution to dealers' optimization problem results in deviations in uncovered interest parity (UIP) condition

$$\mathbb{E}_t[e_{t+1}] - e_t = i_t - i_t^* + \gamma_F \sigma_E^2 (f_t^* + f_{CB,t}^*). \tag{1}$$

Condition (1) determines exchange rate (nominal depreciation) and is called modified UIP condition, because of an endogenous risk premium component. That means, ceteris paribus, increase in capital inflows (f_t^*) or sales of foreign bonds by central bank ($f_{CB,t}^*$) appreciates the exchange rate and vice versa. Magnitude of this effect is influenced by degree of dealers' risk-aversion (γ_F) and volatility of expected depreciation rate (σ_E^2). Capital inflows are deemed to be exogenous and are assumed to evolve according to an AR(1) process.

The last entity comprised in the model is a monetary authority (central bank). Monetary policy of the central bank is presented by means of zero-bounded Taylor-type interest rate rule of the following form

$$i_t = \max\left\{0;\, \rho_I i_{t-1} + (1-\rho_I)\left[\phi_\pi(\pi_t - \pi_{targ,t}) + \phi_P p_t + \phi_Y y_t + \phi_E e_t\right] + \varepsilon_{I,t}\right\}, \tag{2}$$

where ρ_I is interest rate smoothing parameter and $\phi_{(\cdot)}$ correspond to elasticity parameters inflation, price level, output and nominal exchange rate, respectively. Equation (2) introduces zero lower bound constraint for nominal interest rate.

Another constraint to which is the economy exposed can be in accordance with Malovaná [9] expressed by equation

$$f_{CB,t}^* = \min\{0,\, \chi_E(e_t - e_{t-1}) + \chi_{ET}(e_t - e_{targ}) + \chi_Q q_t + \varepsilon_{CB,t}^*\}. \tag{3}$$

Based on values of elasticity parameters $\chi_{(\cdot)}$, central bank can choose from following intervention regimes:

- if $\chi_E = 0$, $\chi_{ET} = 0$ and $\chi_Q = 0$, then central bank performs unanticipated interventions, i.e. $f_{CB,t}^*$ is affected only by a random shock;
- if $\chi_E \neq 0$, then central bank performs managed floating regime, i.e. smooths nominal exchange rate movements by selling foreign bonds to prevent depreciation and vice versa;
- if $\chi_{ET} \neq 0$, then central bank performs fixed targeting (commitment)[1] or
- if $\chi_Q \neq 0$, then central bank follows real exchange rate rule, i.e. corrects a misalignment of the real exchange rate from its long-run level.

3 Methodology

This section follows description of OBC algorithm based on papers Holden [5] and Holden & Paetz [6]. We start with dynamic model under rational expectations in the variables $x_{1,t}, \ldots, x_{n,t}$ and we are studying the response of a particular shock ε_t. When doing so, suppose the model's equations are linear, except one of the form

$$x_{1,t} = \max\{0, \mu_1 + \phi'_{-1} x_{t-1} + \phi'_0 x_t + \phi'_1 \mathbb{E}_t[x_{t+1}] - (\phi'_{-1} + \phi'_0 + \phi'_1)\mu\}, \tag{4}$$

where column vector x_t contains all model variables at time t, vector $\mu = (\mu_1, \ldots, \mu_n)'$ denotes variables' steady state values and $\mu_1 > 0$. Note that a bounded variable in any linear model can be transformed into equation of the same form as in (4).

[1] Note that in the limiting case of $\chi_{ET} \to \infty$, central bank is able to use unlimited amount of interventions while mitigating any deviations of exchange rate from its target.

Now, we introduce shadow price shocks forcing the bounded variable to be higher than expected, when a particular shock drives it to nil for some number of periods. Let v_i be a column vector containing the relative impulse responses of variable x_i in the model without bounds and let $T^* \leq T$ be the number of periods after which the constraint is no more expected to bind.

The occasionally binding constraints (OBC) algorithm starts with replacing equation (4) with

$$x_{1,t} = \mu_1 + \phi'_{-1} x_{t-1} + \phi'_0 x_t + \phi'_1 \mathbb{E}_t[x_{t+1}] - (\phi'_{-1} + \phi'_0 + \phi'_1)\mu + \sum_{s=0}^{T^*-1} \varepsilon^{SP}_{s,t-s}, \tag{5}$$

where $\varepsilon^{SP}_{s,t}$ is a shadow price shock at time horizon s that is known at t, but does not materialize until $t + s$. When calculating impulse responses, all news are realized in time period 0 and the task is to find values for shadow shocks satisfying

$$\varepsilon^{SP}_{s,t} = 0 \qquad \text{for} \qquad t \neq 0. \tag{6}$$

Let $m_{i,s}$ be a column vector comprising relative impulse responses of variable $x_{i,t}$ to the shadow price shock $\varepsilon^{SP}_{s,t}$. Horizontal stacking of responses of corresponding variables creates matrix $M_i = (m_{i,0}, \ldots, m_{i,T^*-1})$. Then, for any vector $a = (a_0, \ldots, a_{T^*-1})'$, impulse responses to a simultaneous shock to ε_t of magnitude 1 and to shadow price shock of magnitude a can be calculated. Total IRF to these shock for variable x_i is equal to $\mu_i + v_i + M_i a$.

The problem is to find value of a, consistent with rational expectations satisfying the constraint. In addition, the shadow price shock at time s can be non-zero only if the bound binds at that time horizon, or equivalently

$$a'(\mu^* + v^* + M^* a) = 0_{T^*}, \tag{7}$$

where $\mu^* = \mu_1 1_{T^*}$, first T^* elements of v_1 are denoted by v^* and M^* is the upper $T^* \times T^*$ submatrix of M_1. This problem can be solved by means of quadratic programming as

$$\widehat{a} = \underset{\substack{a \geq 0_{T^*} \\ \mu^* + v^* + M^* a \geq 0_{T^*}}}{\operatorname{argmin}} \left[a'(\mu^* + v^*) + \frac{1}{2} a'(M^* + M^{*\prime}) a \right] \tag{8}$$

and the sufficient condition for existence and uniqueness of solution to (8) is positive definiteness of matrix $(M^* + M^{*\prime})$. However, there can be multiple or no solutions to (8). In the case of multiple solutions, algorithm chooses a solution with minimal Euclid norm $\|a\|_2$. Contrastingly, in the case of no solution to (8) for sufficiently large shocks, bounds are complementary to one another[2]. A necessary condition for existence of an admissible solution to (8) for arbitrary large shocks is that there exists some $a \geq 0_{T^*}$ satisfying $M^* a \geq 1_{T^*}$.

Obviously, the OBC algorithm described above can be generalized to the cases with multiple bounds. Now, each of the variables $x_{1,t}, \ldots, x_{n^*,t}$ is bounded and for each the equation of type (4) holds. Analogously to approach described for one bounded variable, we add shadow shocks (total of $n^* T^*$ shocks) to each of respective equations and horizontally stack the impulse responses of variable $x_{i,t}$ to each of the shadow shocks in the equation for $x_{k,t}$ into matrix $M_{k,t}$.

Let $M^*_{i,l}$ be the upper $T^* \times T^*$ submatrix of $M_{i,l}$ and let M^* be $(n^* T^*) \times (n^* T^*)$ block matrix with (i, l)-th block $M_{i,l}$ for $i, l = 1 \ldots, n^*$. We again define $\mu^*_l = \mu_l 1_{T^*}$, denote first T^* elements of v_l as v^*_l, and μ^* and v^* are vectors of length $n^* T^*$ block vectors with l-th block μ^*_l and v^*_l, respectively. Analogously to the case of one bounded variable, solution (8) yields the required combination of shadow shocks with existence and uniqueness conditions being identical.

4 Results

In what follows, simulation results from empirical research will be presented. There are always impulse responses for twelve key economic variables depicted. In each figure, fixed targeting (FT), managed floating (MF) and real exchange rate rule (RERR) for both positive (+) as well as negative (-) shocks are considered. Reported figures show logarithmic deviations from respective steady states in percent.

[2] This may happen when hitting one constraint increases the probability of hitting another and vice versa

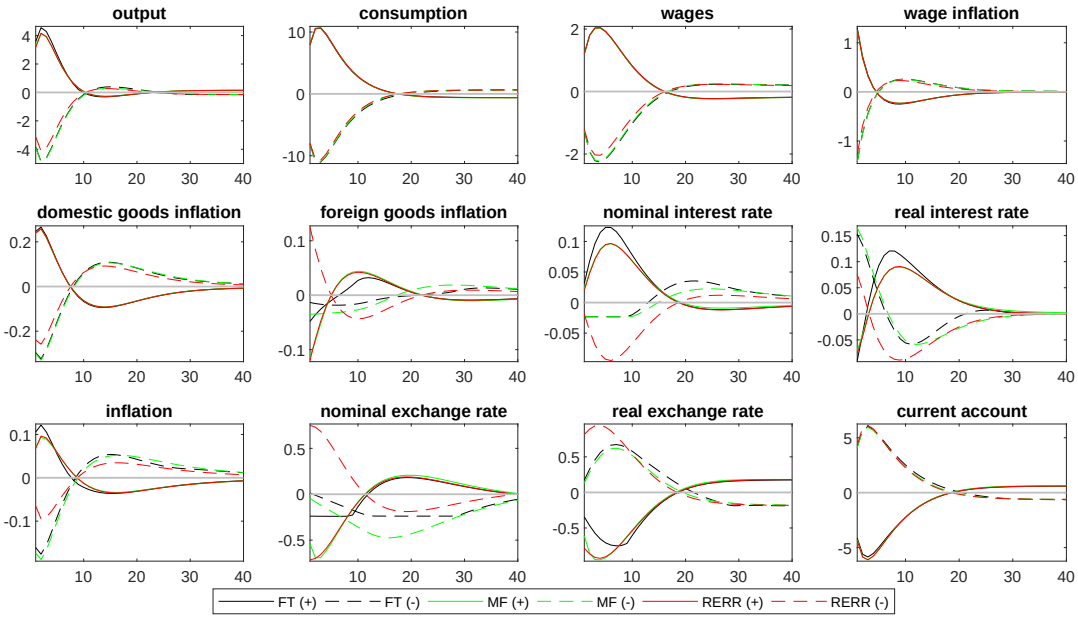


Figure 1 Impulse responses to demand shock.

Impulse responses in a case of domestic demand shock are depicted in Figure 1. Commitment supports the economy at the most in the terms of output gap. This is mainly stemming from reaction of higher interest rates and constrained nominal exchange rate. When negative shock materializes, use of interest rate as monetary policy instrument is limited as it hits the zero lower bound. In this situation, central bank cannot mitigate negative effects sufficiently. Hence, MF or RERR intervention regime could be more appropriate to implement.

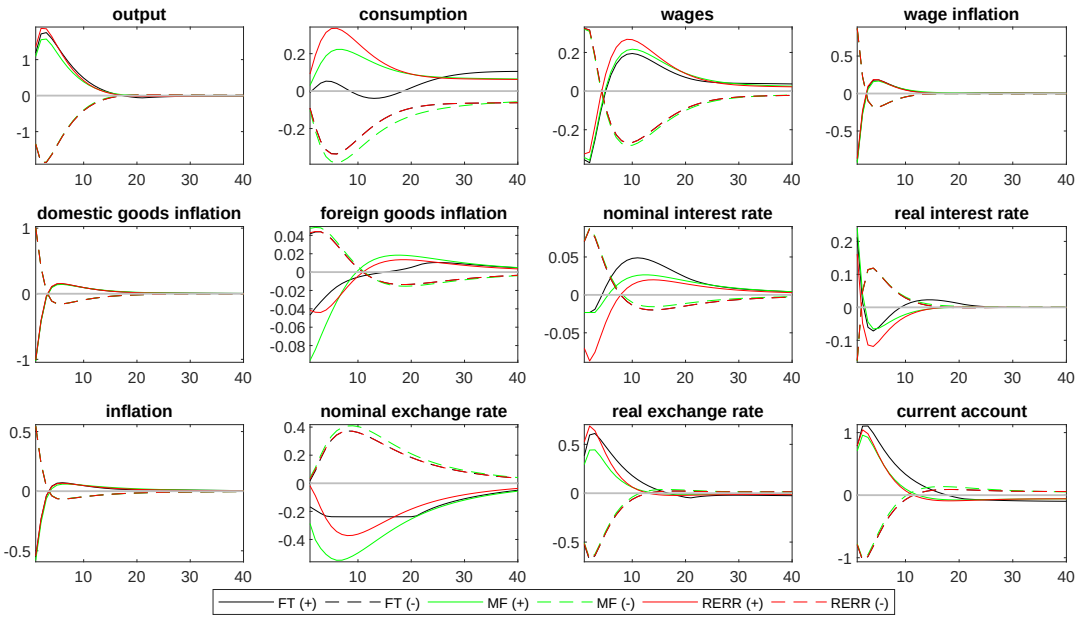


Figure 2 Impulse responses to productivity shock.

When a positive productivity shock hits the economy (Figure 2), higher interest rates force domestic currency to appreciate. Since the currency in a case of commitment hits its bound, this causes relative lower consumption development over time with respect to other intervention regimes. Negative shock is almost invariant to regime chosen by the central bank.

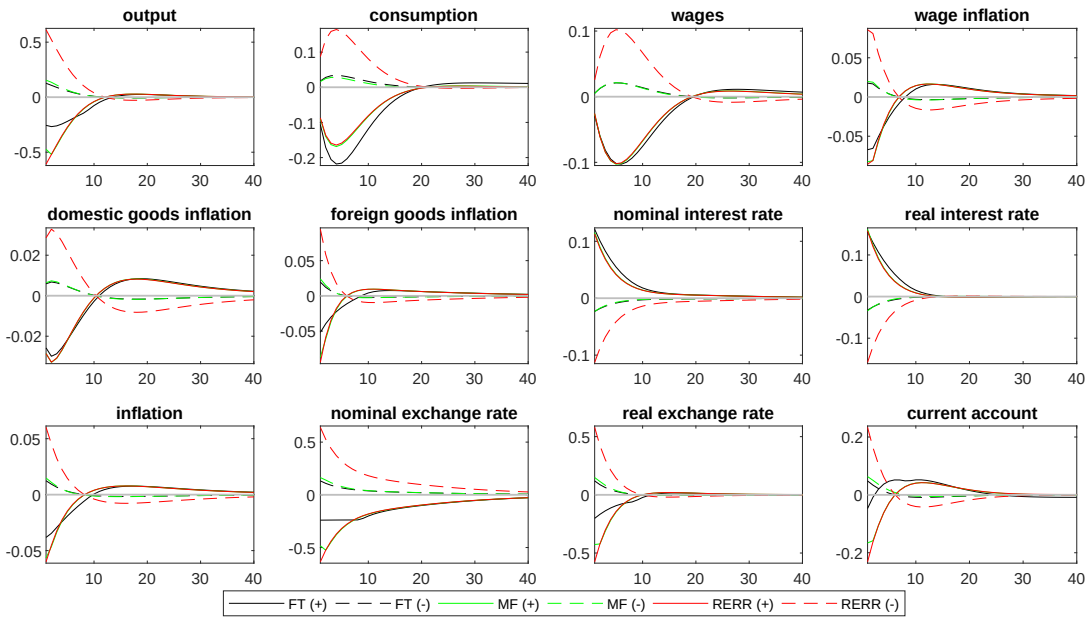


Figure 3 Impulse responses to nominal interest rate shock.

Negative monetary shock in Figure 3 forces nominal interest rate to be affected by zero lower bound constraint. Following commitment or managed floating regime therefore cannot contribute to economic recovery as much as in the case of RERR regime. Conversely, sudden escalation of interest rates causes domestic currency appreciation. Because of hitting the commitment threshold, economy is not exposed to negative consequences as for RERR intervention.

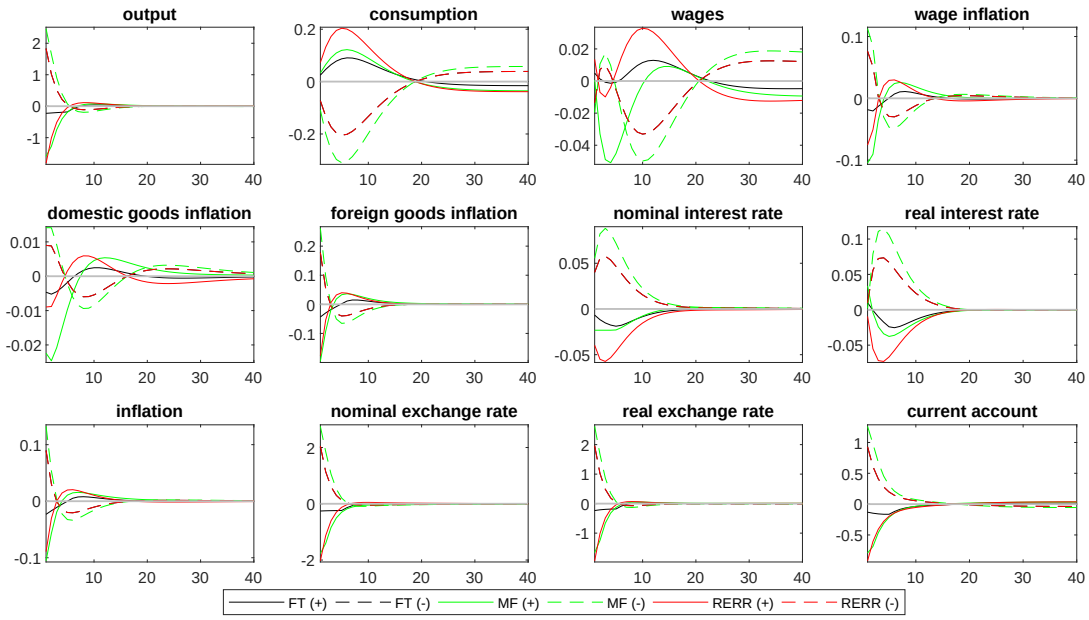


Figure 4 Impulse responses to capital inflow shock.

Unexpected capital outflow (in Figure 4) causes nominal depreciation as well as increase of interest rates in each case, and therefore does not violate any bound. The severity of this shock is higher in a case of MF regime. Unanticipated inflow of the capital considerably reduces volatility of responses to selected variables. In particular, output decreases and deflationary pressures are considerably mitigated.

5 Conclusion

We have inspected possible economic behaviour of key endogenous variables under several intervention regimes set by the central bank by means of impulse response functions. For some of the variables, their development over time is invariant to selected intervention regime. Nevertheless, when there any differences are present, commitment has been proven to be the most successful intervention regime for Czech small open economy with respect to possible mitigation of negative consequences caused by random shock.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0963/2018. This support is gratefully acknowledged.

References

[1] Alpanda, S., Kotzé, K. & Woglom, G. (2010). The role of the exchange rate in a New Keynesian DSGE model for the South African economy. *South African Journal of Economics*, vol. 78(2), pp. 170-191.

[2] Brůha, J. & Tonner, J. (2017). *An Exchange Rate Floor as an Instrument of Monetary Policy: An Ex-post Assessment of the Czech Experience*. CNB Working Paper Series, 4/2017.

[3] Calvo, Guillermo A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, Elsevier, vol. 12(3), pp. 383-398.

[4] Fodor, J. & Vašíček, O. (2017). Unconventional monetary policy in a small open economy under inflation targeting regime. In P. Pražák (Ed.), *35th International Conference Mathematical Methods in Economics MME 2017 Conference Proceedings* (pp. 177-182). Hradec Králové: University of Hradec Králové.

[5] Holden, T. (2017). *Existence, uniqueness and computation of solutions to dynamic models with occasionally binding constraints*. Guildford: University of Surrey. Preprint.

[6] Holden, T. & Paetz, M. (2012). *Efficient simulation of DSGE models with inequality constraints*. Guildford: University of Surrey. Discussion Papers in Economics 16/2012.

[7] Justiniano, A. & Preston, B. (2010). *Monetary policy and uncertainty in an empirical small open economy model*. Working Paper Series WP-09-21, Federal Reserve Bank of Chicago.

[8] Lízal, L. & Schwarz, J. (2013). *Foreign Exchange Interventions as an (Un)conventional Monetary Policy Tool*. BIS Papers No 73. pp. 127-143.

[9] Malovaná, S. (2015). *Foreign Exchange Interventions at the Zero Lower Bound in the Czech Economy: A DSGE Approach*. Prague: Charles University. Working Papers IES 2015/13.

[10] Montoro, C. & Ortiz, M. (2013). *Foreign Exchange Intervention and Monetary Policy Design: A Market Microstructure Analysis*. Basel: Bank for International Settlements. Draft.

[11] Steinbach, M., Mathuloe, P. & Smit, B. (2009). An open economy New Keynesian DSGE model of the South African economy. *South African Journal of Economics*, vol. 77(2), pp. 207-227.

[12] Súkupová, K. & Vašíček, O. (2016). Modelling of an asymmetric foreign exchange rate commitment in the Czech economy: DSGE model with constraints. In A. Kocourek & M. Vavroušek (Eds.), *34th International Conference Mathematical Methods in Economics 2016 Conference Proceedings* (pp. 797–802). Liberec: Technical University of Liberec.

[13] Súkupová, K. & Vašíček, O. (2016). Multiple constraints applied into the DSGE model: The case of unconventional monetary easing in the Czech economy. In M. Reiff & P. Gežík (Eds.), *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII*. (pp. 356–361). Bratislava: University of Economics.

[14] Tvrz, S., Tonner, J. & Vašíček, O. (2018). Behavior of the Czech small open economy in the conditions of occasionally binding constraints. In Váchová, L., Kratochvíl, V.. *36th International Conference Mathematical Methods in Economics Conference Proceedings* (pp. 612-617) Prague: University of Economics.

Factors influencing the yields of the most widespread crops

Tereza Jägerová[1]

Abstract. The crop yield varies considerably between years. Farms influence yields mainly by using fertilizers and pesticides, or by using suitable crop rotation cycles. In addition to factors that are influenced by humans, climate factors, such as air temperature, precipitation and soil moisture - affect crop yield. Extremely warm and dry years have caused significant decrease in harvests in recent years. The regions are better or worse adapted to extreme conditions and therefore the crop yield varies depending on the location of the agricultural land. The estimation of the crop yield model is based on the principle of short panel data, which allows incorporation of the heterogeneity of individual regions of Czech Republic. The aim of the paper is to study the most suitable variables that explain the variability of yields of the most widespread crops - wheat, barley, rape and potato. The model assumes existing persistence of the effects of extreme droughts in recent periods on the current crop yield. The outcome of the study is comparison of the sensitivity of selected crops to exogenous climatic factors and human activity.

Keywords: crop yield, panel data, climatic factors.

JEL Classification: C51, Q15, Q54
AMS Classification: 91B76, 62M10

1 Introduction

Farms annually decide what structure of agricultural production to choose for the upcoming vegetative period. Factors influencing this decision include, in particular, the profitability of individual crops. Of course, the goal is to achieve the highest possible yields so that the profit per hectare of land is as high as possible. However, individual crops exhibit different sensitivity to climatic factors and, depending on the location of the sown land, there is considerable variation in yield. In recent years there have been frequent periods of drought in Czech Republic, which have a devastating effect on crops. Every region adapts differently to unfavorable weather. In general, regions whose land use is highly diversified are less vulnerable than those with large areas of sown land and less forest and grove. In addition, the unfavorable effects of drought add up as the soil dries to a greater depth and it is more difficult for crops to reach the moisture.

An overview of developed models covering in particular the effects of climate and other changes on agricultural production is mapped in [5]. In the world, the impacts of climate change on crop yield are an often discussed topic. The AGMIP model is a major representative of these models in recent years as is shown in [11]. AGMIP calculates the effects of climate change in adapting business behavior to climate change based on different climate scenarios and impacts without adjustment. The model is designed for the whole world and can be calibrated for individual regions. So far, an application of AGMIP to South Asia and Sub-Saharan Africa has been published [10]. One of the important works regarding wheat yield uncertainty is [1]. In the US, a growth model CLM-APSIM was developed at the University of Illinois to investigate the impact of climate change on the profitability of some crops [8]. The relationship between climate factors and crop yield in the Czech environment discuss for example [15].

Crop yield prediction models are used either for predictions within the current period (growth models for yield estimates of already growing crops), for next period yield estimates, or for assessing climate scenarios over several years (some complex models have been mentioned above). The crop yield model itself is modeled mainly by a multiple linear regression [4, 7, 9, 12], or by Cobb-Douglas production function [9] or in recent years by adaptive neuro-fuzzy modeling [9] and data mining methods (neural networks [12], clustering [6], random forest [4] and others tree methods [2] are used to explain the variation in crop yield, especially wheat and maize). These models are often compared to econometric models and often achieve better results. However, the disadvantage of the data-mining methods is that the effect of individual variables on yield is not easy to interpret and, in addition, a large amount of data is needed to train the model. An overview of the methods used

[1] University of Economics, Prague, tereza.jager@gmail.com.

to model crop yield predictions is given in [7]. Some of the mentioned models deal with the analysis of factors of production (labour, fertilizers, pesticides, water, electricity, diesel, ...) on the yield for the study of their best combination [9], another group of models represents growth models within the current season [2, 11] and the last group analyzes the effects of climate factors [4]. The random forest model [4] uses the average climate variable in the current period, the level of irrigation and the use of nitrogen-based fertilizers to predict crop yield. It compares the use of random forest method and multiple linear regression, which shows worse results under given conditions. This model is closest to the crop yield prediction models below. However, the random forest model does not anticipate the persistent effects of long-term drought and high temperatures and is rather intended to model long-term forecasts of wheat, potato and corn yields. Most models are not designed to plan the production structure for a subsequent period.

The aim of the paper is to study climatic factors and human activities, which have the most significant impact on yield of the most widespread crops – wheat, barley, rape and potato. The model assumes existing persistence of the effects of extreme droughts in recent periods on the current crop yield. Human activities include the use of fertilizers and pesticides; the study abstracts the crop rotation cycles and other factors influenced by humans. The influence of heterogeneity of individual regions of the Czech Republic is included in the model using Panel data. Based on the identified factors, a crop yield prediction models were constructed. The model assumes the use of at least seven-month weather forecasts published by the European Center for Medium-Range Weather Forecasts and is intended to support farmers' decision-making.

2 Data

Public data for regional crop yields and climate variables have been available since 2002. Data on crop yields within Czech regions in time series is provided by the Czech Statistical Office. Monthly averages of temperatures and precipitation and their long-term values (1981 – 2010) are available on the website of the Czech Hydrometeorological Institute. Quantitative levels of pesticides and fertilizers used per hectare of crop production are not publicly available. To incorporate the effects of used fertilizers and pesticides, hectare costs for each crop were used as proxy variables. The cost data on fertilizers and pesticides by crops are provided by the Institute of Agricultural Economics and Information as part of annual agricultural crop cost surveys. The time series of industrial producer price indices is published by the Czech Statistical Office. The data for the models below correspond to the time series from 2002 to 2017.

3 Methodology

Climate variables were used to explain the variability of crop yields - average monthly temperatures and rainfall from March to August. To clarify the abbreviations used for each variable, a list is given below:

- T3 – T8 the monthly temperature from March to August;
- R3 – R8 the monthly rainfall from March to August.

To include the potential impact of past periods on the current yield, we construct aggregated temperature and precipitation indicators. The aim is to create such indicators that enhance the influence of significant variations in climate variables from their long-term normal values. On the other hand, the inverse value of the indicator (1) is used to limit the influence of extremely high precipitation (several times more than the long-term average), which caused extensive floods especially in 2002 and 2010. The indicators are constructed on the basis of climate variables in the past vegetative period, i.e. from September to August. Aggregated indicators are integrated into models using their actual and delayed values.

$$D_i = \frac{\sum_{m=1}^{M} N_m^2}{\sum_{m=1}^{M} CV_{m,i}^2}, \qquad m = 1,2,\dots,M, \qquad i = 1,2,\dots,T. \tag{1}$$

Where: N_m is long-term average of climate variable in month m and $CV_{m,i}$ is value of climate variable in month m of year i.

Let denote the aggregated temperature indicator D1 and the aggregated rainfall indicator D2. The mean and variance of D1 are 0.9073 and 0.005665; the same statistics of D2 are 0.8924 and 0.06881 respectively. An average of D1 below 1 indicates temperatures above their normal values over the long term, as shown in Figure 1 (TV is temperature variation in degrees Celsius). On the contrary, the average value of the indicator D2 lower than 1 does not mean higher volume of precipitation than the normal value in the long term, but the existence of

a large number of extremely high precipitation and long droughts. The increased values of extremely high precipitation cause an impression of an excess of total rainfall. The distribution of variations of monthly precipitation values in relation to the corresponding long-term normal values (RV = 100) is shown in Figure 2 (RV denounces rainfall variation in percent).

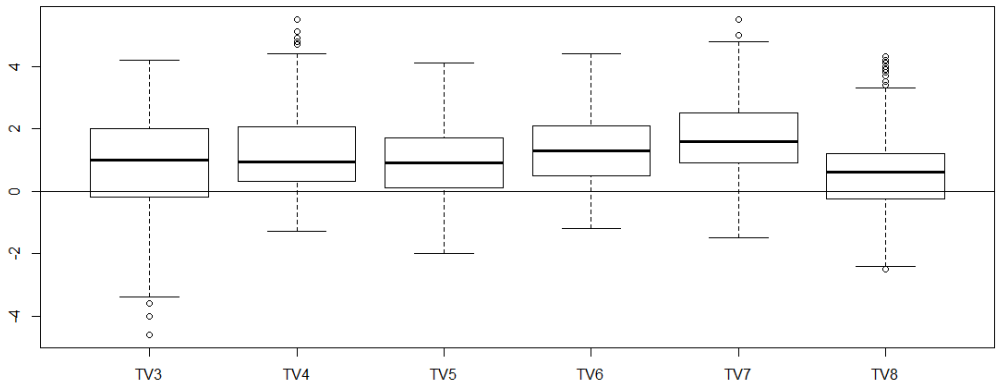


Figure 1 Variations in average monthly temperatures from long-term normal values from March to August

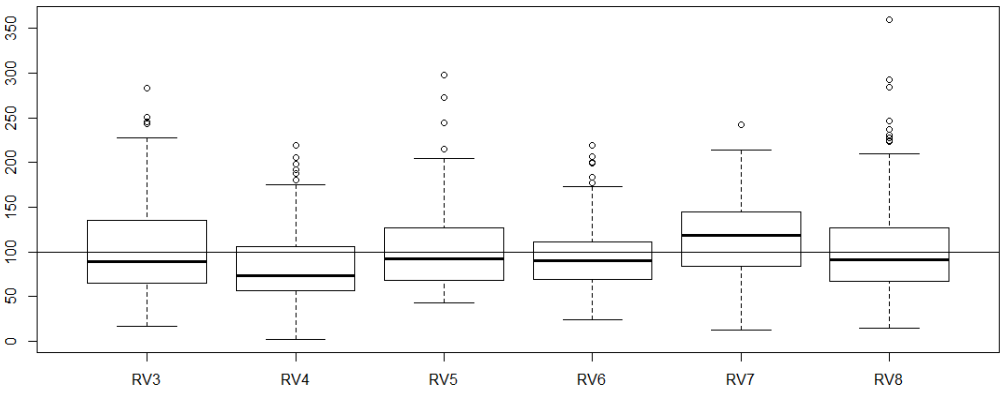


Figure 2 Variations in average monthly precipitation from long-term normal values from March to August

Figure 2 shows that, apart from the precipitation in July, all average values are below long-term normal values and with a relatively high number of extreme rainfall. Values of indicator D2 less than 1 may mean two situations: rainfall was generally above average in a given year or extreme weather fluctuations (long droughts alternating with extremely high rainfall). It follows from the data of the monitored period that in the last years we encounter rather the second possibility. This situation causes the sum of the squares of the precipitation variables in the reference period to be greater than the sum of the squares of the long-term normal precipitation values due to squaring highly above-average precipitation values and the D2 indicator to be less than 1.

Panel data allows capturing the differences between regions of the Czech Republic. Individual regions show considerable differences in their character. They differ, for example, in the quality of the soil, the altitude or the segmentation of the landscape - the prevailing flat or slightly wavy terrain or mountain area. The impact of these strictly exogenous factors is difficult to measure, or it is impossible because of the unavailability of data. These factors represent the unobserved region heterogeneity, which is unchanged over time. The time series of data available to estimate crop yields correspond to short time series (annual data 2002 to 2017). Due to the length of time series and the existence of unchanged heterogeneity between regions, methods for so-called Short panels can be used. The crop yield time series do not show a time trend according to the values of the extended Dickey-Fuller test ADF [14] (ADF for wheat is –6.32, barley –6.92, rape –8.45 and for potato –4.39, all below critical value –3.42). The basic forms of individual effects were used, fixed and random. Wooldridge [13] discusses

Panel data in detail. The suitability of using fixed effects over a classical linear regression model was tested by the F-test recommended in [13]. The Hausman test (for example in [13]) verifies the adequacy of using random effects over fixed ones. To estimate models with individual effects, we analyze the variability of yields due to changes in time within regions rather than between regions. We do not expect crop yields to be affected by the situation in neighboring regions.

Financial variables (pesticide and fertilizer costs) were adjusted for inflation using industrial producer price indices (CE 20 for pesticides and CE 201 for fertilizers). The influence of exogenous technological progress is integrated into the models either in a linear or quadratic expression of the time trend.

The suitability of the combination of explanatory variables in the model was assessed through the development of the Akaike information criterion AIC [3], the coefficient of determination R^2 and the predictive ability of the model evaluated by the MAPE indicator (i.e., the percentage deviation of the predicted values from the actual). Due to the unavailability of theoretical assumptions as to the appropriateness of the explanatory variables, they were selected by the reverse step method. The starting point of the analysis was a model with all variables from the set of relevant explanatory variables (see above). The variables were subsequently collected according to the smallest increase of the AIC value, until the given combination of variables and possible integration of the individual effects of the regions showed the best predictive results (according to MAPE value) at a reasonable explanation level of the yield variability by the model (coefficient of determination R^2). The inclusion of pesticide costs and fertilizer costs in one model caused collinearity (tested by Variance Inflation Factor – e.g. in [14]). The cost variable that contributed less to the yield explanation has been removed.

The model was estimated on time series data between 2002 and 2015. The last 2 years of time series were used to verify the accuracy of the estimated models.

4 Crop yield models

For each crop, the suitability of explanatory variables was analyzed using the Akaike information criterion, the MAPE indicator, and the coefficient of determination R^2. These variables were used to estimate the crop yield model. An appropriate form of individual effects was tested using the F-test and the Hausman test. Table 1 shows the results of the estimated yield models – the estimated parameters of the explanatory variables of each model, the coefficient of determination R^2 and the accuracy of the forecasts (MAPE values for the whole verification period and also for the years 2016 and 2017 separately). Robust standard errors are in parentheses below model parameter estimates (corresponding levels of significance: * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$).

Estimated models show a large degree of model-data matching and high prediction accuracy (especially wheat and barley yield model). All models were verified on data from 2016 and 2017. The variations in forecasts for all models differed from reality by 3.70% on average. Wheat, barley and potato yield model have the fixed character of individual regional effects. The differences between the individual effects of the regions in the rape yield model were not statistically significant.

The increasing average temperature from May to August has a negative impact on crop yields (the exception is the effect of rising temperatures in July, i.e. just before harvest, on the yield of rape). Warmer April also has a positive impact on rape yield. The impact of monthly rainfall on crop yields was statistically significant only for June and July.

The effect of the evolution of monthly temperatures in the past has amplified or diminished the effect of precipitation volume over the same period. High temperatures with low rainfall cause significant soil drying and this phenomenon due to its persistence affects soil fertility even in the subsequent vegetative period. Individual crops are differently sensitive to changing soil moisture, and this effect is not statistically significant for all crops. The effect is best seen in the estimated potato yield model - lower temperatures over the past period (higher D1 with a positive coefficient) together with lower temperatures and continuous rainfall (higher D2 with a positive coefficient) in the current vegetative period will cause a significant increase in hectare yield. Negative estimation of the D1 coefficient in the current period moderates the negative impact of high temperatures on yield.

The estimated parameters for indicator D2 in the yield functions take the positive values, which imply a possible relationship - the more extreme values in the current or previous vegetative period, the lower the yield of the crop. As a result, it is important for crops to avoid excessive fluctuations in the volume of precipitation. A more pleasant situation is also the smaller volume of rainfall but continuously dispersed throughout the year. This leads to a better irrigation of the soil from which the water does not flow off without benefit. In the case of the delayed value of the indicator D2, the interpretation is similar. Continuous rainfall results in better soil irriga-

tion, which conserves some moisture to the next vegetative period and does not dry the soil, which has a negative impact on yield.

The rising costs of used pesticides and fertilizers are expected to have a positive effect on yield. Exogenous technological progress is included in the wheat and potato yield models using a linear trend; in the barley and rape yield models, the quadratic trend was more suitable.

	wheat	**barley**	**rape**	**potato**
Intercept	Fixed effects	Fixed effects	6.0757 (0.4781)***	Fixed effects
T4			0.1600 (0.01302)***	
T5	–0.3234 (0.01463)***	–0.1814 (0.01609)***		
T6			–0.2830 (0.01997)***	
T7	–0.2119 (0.01898)***	–0.1063 (0.006715)***	0.1330 (0.01348)***	–2.0123 (0.1597)***
T8				–1.7799 (0.1334)***
R6	–0.006172 (0.0007856)***			
R7	–0.004714 (0.0008376)***			0.01874 (0.004105)***
$D1_t$		2.3190 (0.1758)***		–8.9768 (2.2953)***
$D1_{t-1}$	–2.6263 (0.3262)***	–2.0797 (0.1355)***	–2.2425 (0.22650)***	7.6928 (1.7285)***
$D2_t$	–0.3664 (0.09445)***	0.3501 (0.02317)***		2.2106 (0.3598)***
$D2_{t-1}$			0.4834 (0.08162)***	
Pesticides				0.0005588 (0.0001470)***
Fertilizers	0.0004420 (0.00002676)***	0.0002528 (0.00002892)***	0.0001652 (0.00002587)***	
Exogenous technological progress	0.06711 (0.005579)***	0.005614 (0.0002743)***	0.003740 (0.0003434)***	0.5919 (0.04137)***
R^2	**0.8225**	**0.9076**	**0.8339**	**0.8174**
MAPE	**2.8791**	**2.8323**	**4.7704**	**4.2469**
$MAPE_{2016}$	2.0450	1.7124	6.3312	4.9565
$MAPE_{2017}$	3.7132	3.9521	3.2096	3.5374

Table 1 Estimation of crop yield models

The South Moravian Region is the best locality to cultivate wheat considering maximal value of its fixed effect $FE_W = 15.51$. On the contrary, the Karlovy Vary Region is the least suitable for it on average ($FE_W = 13.86$); barley is also doing the worst here ($FE_B = 6.47$). The ideal location for barley is the Olomouc Region ($FE_B = 7.54$) and the Pardubice Region for potatoes ($FE_P = 83.45$). In the Ústí Region, potatoes are doing the least ($FE_P = 79.93$). The largest relative differences between the fixed individual effects of the regions are in the wheat yield model, a maximum is 14 percent from the average yield (differences related to the average yield level of the crop). The variability of regional effects in other models ranges up to 8 percent from the average. The fixed effect values for individual regions correspond to the generally known data on the fertility of production areas in the Czech Republic and confirm the functionality of the estimated models.

5 Conclusion

This paper investigated the factors that influence the yield of the most widespread agricultural crops - wheat, barley, rape and potato. Empirical analysis used short panel data modeling techniques; wheat, barley and potato yield models correspond to fixed individual effects. The impact of the different character of Czech regions on crop yields is most noticeable in wheat yield model. Conversely, the differences between regions in the rape yield are small, so the most appropriate is pooled model without individual regional effects. The regional fixed effects correspond to the generally known data on the fertility of production areas in the Czech Republic and confirm the functionality of the estimated models. The results show significant negative persistent impacts of extreme climatic conditions (long droughts and then extremely high rainfall) during the previous vegetative period on crop yields in the current period. These phenomena were modeled using aggregated indicators of monthly temperatures and precipitation deviations from their normal values for each Czech region. The combination of these variables contributes to explain the impact of past climatic conditions on crop yields. In addition to the effects of delayed climatic factors, the effect of monthly temperature and precipitation variables in the current period was also analyzed (in their isolated and aggregated form). In general, the course of last three months before harvest has a major impact on crop yields together with the continuing effects of possible heat and drought periods from the past and current periods. The costs of used pesticides and fertilizers for individual crops have had a positive effect on crop yields as expected. All models were validated on 2016 and 2017 data and show high and stable prediction accuracy.

Acknowledgements

Great thanks go to Professor Jan Pelikan from the University of Economics in Prague for his guidance on research methods and ideas.

References

[1] Asseng, S. et al. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, *3*, 827–832.
[2] Cai, Y. et al. (2017). Crop yield predictions - high resolution statistical model for intra-season forecasts applied to corn in the US. *American Geophysical Union, Fall Meeting 2017*.
[3] Cavanaugh, J. E. (1997). Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters, 31*, 201–208.
[4] Jeong, J. H. et al. (2016).Random Forests for Global and Regional Crop Yield Predictions.*PLoS ONE,11(6)*.
[5] Jones, J. W. (2016). Brief history of agricultural systems modeling. *Agricultural Systems, 155*, 240-254.
[6] Manjula, E. & Djodiltachoumy, S. (2017). A Model for Prediction of Crop Yield. *Internation Journal of Computational Intelligence and Informatics, 6 (4)*, 298-305.
[7] Menaka, K. & Yuvaraj, N. (2016). A survey on crop yield prediction models. *Indian Journal of Innovations and Developments, 5 (12)*.
[8] Peng, B. et al. (2018). Improving maize growth processes in the community land model: Implementation and evaluation. *Agricultural and Forest Meteorology*, *250–251*, 64–89.
[9] Qaddoum, K. et al. (2011). Adaptive neuro-fuzzy modeling for crop yield prediction. *Recent Researches in Artificial Intelligence, Knowledge Engineering and Databases, 16(7)*, 1-6.
[10] Rosenweig, C. et al. (2012). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Integrated regional assessment projects. In: Hillel, D.,Rosenzweig, C. (Eds.), Handbook of Climate Change and Agroecosystems: Globaland Regional Aspects and Implications. *ICP Series on Climate Change Impacts, Adaptation, and Mitigation Vol. 2.*, Imperial College Press, 263–280.
[11] Rosenweig, C. et al. (2013). The Agricultural Model Intercomparison and Improvement project (AgMIP): Protocol and pilot studies. *Agricultural and Forest Meteorology, 170*, 166–182.
[12] Shastry, K. A. et al. (2016). A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction. *Journal of Artificial Intelligence, 9*, 23-32.
[13] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data, 2E.* London: MIT Press.
[14] Wooldridge, J. M. (2012). *Introductory Econometrics: a modern approach 2E.* South-Western Cengage Learning.
[15] Žalud, Z. (2009). *Změna klimatu a české zemědělství - dopady a adaptace.* Brno: Mendelova zemědělská a lesnická univerzita.

An Empirical Analysis of Macroeconomic and Bank Performance Factors Affecting Credit Risk in Banking for The Central European Countries

Xiaoshan FENG [1]

Abstract. This paper provides empirical results of the macroeconomic and bank performance determinants which affect credit risk in banking for the Central European Countries. Using time series data from 2002 to 2016 for Austria, the Czech Republic, Germany, Hungary and Poland. In this study, we apply fixed effect panel data regression to figure out how banks' non-performing loans are influenced by macroeconomic factors and bank performance indicators.
Through econometric verification, the valid estimation shows empirical results that macroeconomic and bank performance factors do have significant impact on credit risk. GDP growth rate, share price indices, and return on assets are negatively related to the increasing of NPL ratio, while better banks' assets quality is following with lower exchange rate, harmonized consumer price index and unemployment rate.

Keywords: Econometric analysis, Panel data regression, Credit risk, Non-performing loans, Macroeconomic variables, Bank performance indicator, Central Europe.

JEL Classification: E44, C33, C52,
AMS Classification: 91G40 credit risk , 91G70 Statistical methods, econometrics

1 Introduction

For most of banks, the major risk is credit risk. Generally, credit risk is associated with the traditional lending activity of banks and it is simply described as the risk of a loan not being repaid in part or in full. The non-performing loan ratio of a bank is a key indicator to measure credit risk management, it shows the quality of banks' assets. Moreover, NPL ratio not only has a significant impact on the daily operation and development of banks, but it also plays a decisive role in the financial stability of a country. When economic in recession, default probability will rise due to several possible factors, some of the factors cannot be controlled via bank's internal management, such as macroeconomic factors, therefore, to understand the significant factors which may lead credit risk worse could be a better way to react for regulatory authority.

The aim of this paper is to evaluate how credit risk is influenced by macroeconomic and bank performance determinants for selected countries in Central Europe. We choose five Central European countries with rapid GDP growth rate, i.e. Austria; the Czech Republic; Germany; Hungary; Poland. Time series range from 2002 to 2016, data collected from the database such as OECD and World Bank. This paper is divided into five sections. The first section starts with introduction and the last one ends with the conclusion. The second section includes the literature review, the third one is description of methodology and data collection. In the fourth section, empirical result will be discussed.

2 Literature review

In this section, we will provide a review of the previous studies concentrated in effects of macroeconomic and bank-specified factors on bank's credit risk.

There is rich literature examine the significance of macroeconomic factors on credit risk of banking. Beck, Jakubik and Piloiu [1] has examined from 75 countries during the past 10 years that a decline of stock prices can negatively affect bank assets quality, in particular in countries with large stock markets relative to GDP. Therefore, this study concludes that for countries with small stock markets capitalization relative to GDP, the effect is not statistically significance.

[1] VŠB – Technical University of Ostrava, Department of Finance, Sokolská třída 33, 702 00 Ostrava, xiaoshan.feng@vsb.cz

Jakubík et al. [5] have investigated credit risk modeling with macroeconomic variables based on the different sectors from 1998-2006 for the Czech and the German economy. For corporate sector, real GDP growth rate and inflation rate have negative impact on non-performing loan ratio, while the real exchange rate positively affects the firms' default rate. As for private sector, unemployment rate and real interest rate have significant impacts on non-performing loan ratio negatively.

Louzis, Vouldis, and Metaxas [6] used dynamic panel data methods to examine the determinants of non-performing loans in the Greek banking sector from 2003-2009, separately for consumer loans, business loans and mortgages. The study is based on the hypothesis that both macroeconomic and bank-specific variables influence the loan quality, while the effects vary between different loan categories. The results show that no matter what kind of loan categories are, macroeconomic variables such as GDP, unemployment, interest rate has strong impact on the level of NPLs. In particular, consumer loans are the most sensitive to changes in the lending rates and business loans to the real GDP growth rate, while mortgages are the least affected by macroeconomic factors.

Messai and Jouini [8] applied panel data study on 85 banks in Italy, Greece and Spain from 2003-2008, which evaluates microeconomic and macroeconomic determinants of non-performing loans. The results show the profitability of banks' assets and growth rate of GDP negatively affect the NPLs while loans vary positively with unemployment rate, real interest rate and loan loss reserves.

Škarica [9] investigated determinants of non-performing loans in Central and Eastern European Countries (Bulgaria, Croatia, Czech Republic, Hungary, Latvia, Romania and Slovakia). The analysis used fixed effects estimator of panel data regression for the selected European emerging markets from 2007 to 2012. The results show that the real GDP growth was the significant driver of the increase of the NPL ratio during the past 5 years in CEE countries. while inflation and unemployment negatively affect assets quality.

Based on previous research results, although there has much literature investigated the factors affecting the NPL ratio, most of all focused on the macroscopic perspective, while the literature on the impact of the micro factors on the NPL ratio is relatively rare. Therefore, this paper refers to the banks' profitability indicators, from the micro point of view, combined with macroeconomic variables to explore the impact of external and internal factors on the non-performing loans of banks in Central European countries.

3 Methodology and Data Description

In this paper, we are using the panel data sets, which consists of a time series for each cross-sectional member in the dataset. The panel data set consists five countries (Austria; Czech Republic; Germany; Poland; Hungary;) for same period of fifteen years (2002-2016), which has relatively high GDP growth rate among European Union.

The original model can be constructed as follows:

$$Y_{i,t} = \beta_0 + X'_{i,t}\beta + \epsilon_{i,t}\,, i = 1, \dots, N, t = 1, \dots, T. \quad (1)$$

Where the subscripts i and t denote the cross-sectional and time dimension of the panel sample respectively. There is k $(k = 1, \dots, K)$ regressor in $X_{i,t}$, not including a constant term. $X_{i,t}$ is explanatory variable value for i-th section at t-th dimension; β_0 is the intercept; β is the slope coefficient of a $(k \times 1)$ vector. The variable $\epsilon_{i,t}$, can be called as the error term in the relationship, represents factors other than explanatory variables that affect dependent variables.

After we set econometrics model, we can estimate the model and find out the result of the value of slope coefficient β for explanatory variables. While we still need to verify the model to ensure the regression model correctly specified, then promote more good use in the future. Basically, there are four verifications, based on the classical assumptions of model, which are No autocorrelation; No multicollinearity; Homoscedasticity; error term normality.

Autocorrelation is correlation between members of series of observations ordered in time. In the regression the classical linear regression model assumes that such autocorrelation does not exhibit in the error term $u_{i,t}$. The presence of correlation in error term $u_{i,t}$ drastically reduces model's accuracy. The main method for detection is Durbin-Watson (DW) statistic. Although DW test can be used in panel model, we should take one of the limitations for DW test into consideration, if the model contains lagged variables, DW test cannot use to detect autocorrelation, therefore, the another way to detect autocorrelation of residual for panel model is using Pesaran Cross-sectional Dependence (CD) test or Breusch Pegan Lagrange Multiplier (LM)test, the general null hypothesis of no cross-section dependence may be stated in terms of the correlations between the disturbances in different cross-section units:

H_0: $\rho_{ij} = corr(u_{i,t}, u_{j,t}) = 0,\ for\ i \neq j$

H_A: $\rho_{ij} \neq 0$

If we fail to reject H_0, we need to eliminate autocorrelation for more precise prediction. There are various methods for elimination of serial autocorrelation. The most frequently use is to include lagged dependent variable as explanatory variable.

The second verification is to test if the model exhibits *multicollinearity*. There should no perfect multicollinearity among the explanatory variables based on the assumptions underlying the method of least squares. When the explanatory variables are found to be highly correlated, it is hard to figure out the true relationship of a predictor with response variable. To test if our model exhibits multicollinearity, we can compare the squares of coefficients of multi-correlation and coefficient of determination. Compared with the original coefficient of determination which is R-squared. R-squared how well the model fits the data, and the F-test is related to it. For each X_i $(i = 2, \dots, k)$ we estimate this variable using other explanatory variables and we set $R^2_{X_i}$.The hypothesis is:

H_0: $R^2_{X_i} \leq R^2$ (No multicollinearity)

H_A: $R^2_{X_i} > R^2$ (Multicollinearity)

$$F_{cal} = \frac{R^2_{X_i}/(k-2)}{(1-R^2_{X_i})/(k-2)} \sim F(k-2, n-k+1). \quad (2)$$

Where k is number of parameters in model, n is the number of observations.

Thirdly, linear regression analysis requires that there is *homoscedasticity*, in other words, we expect the model has no heteroscedasticity. Which is that the variance of error term $u_{i,t}$ is some constant number. Homoscedasticity means equal variance of error term, while heteroscedasticity means unequal variance. If we still estimate despite heteroscedasticity, the estimation could lead conclusion being misleading. We can use the Breusch-Pagan test to detect if heteroscedasticity exhibits. It usually applied by assuming that heteroskedasticity may be a linear function of all the independent variables in the model, we can regress the squared residuals on explanatory variables in the model, as follow:

$$\widehat{u_{i,t}}^2 = \alpha_1 + \alpha_2 x_{i1} + \alpha_3 x_{i2} + \cdots + \alpha_m x_{im} + \epsilon_{i,t} \quad (3)$$

The null hypothesis is:

H_0: $\alpha_2 = 0 \wedge \alpha_3 = 0 \wedge \alpha_4 = 0 \wedge \alpha_5 = 0 \wedge \alpha_6 = 0$ (Homoscedasticity)

H_A: $\alpha_2 \neq 0 \vee \alpha_3 \neq 0 \vee \alpha_4 \neq 0 \vee \alpha_5 \neq 0 \vee \alpha_6 \neq 0$ (Heteroscedasticity)

From this auxiliary regression, we obtain the R-squared value, then used to calculate the F-statistic:

$$\mathrm{F}_{cal} = \frac{R^2_{\widehat{u_{i,t}}^2}/1}{1-R^2_{\widehat{u_{i,t}}^2}/n-2} \quad (4)$$

If we detect our model has heteroscedasticity, we need to eliminate heteroscedasticity by applying Method of Weighted Least Squares.

Finally, we need to verify the error term is *normally distributed*. If the error terms are non- normally distributed, confidence intervals may become too wide or narrow. Once confidence interval becomes unstable, it may lead to difficulty in estimating coefficients based on minimization of least squares. Exhibits of non-normal distribution suggests that there are a few unusual data points which must be studied closely to make a better model. This assumption can be checked with Jarque-Bera test. The null hypothesis is H_0: residuals are normally distributed.

$$\mathrm{JB} = \mathrm{n}\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right], \quad (5)$$

where S is skewness and K is kurtosis. A tiny p-value and a large chi-square value from this test mean that we can reject the null hypothesis that the data is normally distributed.

In this analysis, we will discuss the relationship between non-performing loan and macroeconomic and bank performance factors. Table 1 shows the data source for our analysis.

Indicators	Units	Abbreviation	Frequency	Database
Non-performing loans ratio	%	NPL	Annual	World Bank
Return on assets	%	ROA	Annual	World Bank
Net interest margin	%	NIM	Annual	World Bank
GDP growth rate	%	GDPG	Annual	World Bank
Harmonized unemployment rate	%	UR	Annual	OECD
Share price indices (2015=100)	%	SPI	Annual	OECD
Consumer price index (2015=100)	%	CPI	Annual	OECD
Exchange rates (national currency per US dollar)	%	ER	Annual	OECD

Table 1 Data description

4 Empirical results and discussion

In this section, we will verify the estimated model then interpret the result of regression, compared with previous literature and theoretical background. In order to test stationarity of observations, we use Unit Root Test in EViews, then find the lagged or difference for variables.

The transformed model can be estimated as follow:

$$NPL_{i,t} = \beta_0 + \beta_1 gdpg_{i,t-1} + \beta_2 dcpi_{i,t} + \beta_3 dspi_{i,t} + \beta_4 dur_{i,t-1} + \beta_5 nim_{i,t} + \beta_6 droa_{i,t} + \beta_7 der_{i,t} + \epsilon_{i,t}, i = 1, \dots, N, t = 1, \dots, T. \quad (6)$$

Therefore, we are looking for explanatory variables including lagging factor such as GDP growth rate in one-year time lag and absolute level of change in factor such as CPI, SPI, etc.

	Diagnostic test	Result	Decision	Eliminate
Autocorrelation	Breusch Pegan Lagrange Multiplier (LM) test	43.0312 (0.0000)	Autocorrelation	Lagged dependent variable
Multicollinearity	$R^2_{X_i}$.v.s. R^2.	Fail to reject H_0	No multicollinearity	
Homoscedasticity	Breusch-Pagan test	17.3528 (0.0000)	Heteroscedasticity	Cross-section weight
Error term normality	Jarque-Bera test	3.4010 (0.18)	Normally distributed	
Note: Significant at 5%				

Table 2 Validation of model

Table 2 shows the brief progress of validation of the model, from which we can generate the model as follow:

$$NPL_{i,t} = \beta_0 + \beta_1 gdpg_{i,t-1} + \beta_2 dcpi_{i,t} + \beta_3 dspi_{i,t} + \beta_4 dur_{i,t-1} + \beta_5 nim_{i,t} + \beta_6 droa_{i,t} + \beta_7 der_{i,t} + \beta_8 NPL_{i,t-1} + \epsilon_{i,t} \quad (7)$$

Then we can interpret the coefficient of the variables based on regression result thus to discuss relationship between NPL and macroeconomic as well as bank performance factors.

Variable	Expected sign	Coefficient	Std. Error	t-Statistic	Prob.
C		1.619603	0.475645	3.405068	0.0013
GDPG (-1)	-	-0.082801	0.036183	-2.288392	0.0262
DCPI	+	0.077302	0.079226	0.975707	0.3337
DSPI	-	-0.010193	0.003927	-2.595339	0.0123
DUR (-1)	+	0.172541	0.101715	1.696321	0.0958
NIM	-	-0.128030	0.207769	-0.616216	0.5404
DROA	-	-0.152476	0.045346	-3.362473	0.0015
DER	+	0.074797	0.022183	3.371754	0.0014
NPL (-1)	+	0.714879	0.038646	18.49802	0.0000
Observations	65		R-squared		0.9456
Effect specification			Cross-section fixed		

Table 3 Regression result of new model

In the estimation, we specified cross-section fixed effect, in which we allow for each cross-section to have a particular effect, the intercept estimators in fixed effect are of special interest they can be used to analyze the extent of variable heterogeneity and to examine any particular cross section that may be of interest.

Table 3 shows the result of new model, R-squared is 0.9456, which indicates there is about 94% of the total variation NPL ratio explained by the regression model, from point of view, this model is quite nice. The constant 1.6196 represent the average effect of all five cross-sections.

The estimates indicate that the GDP growth rate in period t-1 is negatively related to NPL ratio, in other words, 1 percent higher GDP growth rate will lead to NPL ratio decrease by 8 percent. P-value at 0.0262 means the result is statistically significance. The result is strongly supporting to the previous literature, similar to Messai and Jouini [8] and Jakubík et al. [5]. Well economic situation of countries, more credits of corporate, individuals less bad debt for commercial banks, while in the economic recession, the NPL ratio will increase.

The estimates of CPI in first difference, which can be seen as level of change to last period, which is positively related to NPL ratio, similar to Škarica [9]. With respect to the theory, European Central Bank has always aimed to maintain price stability through target inflation, the result in this study shows statistical insignificance, inflation still has a certain stimulating effect on economic growth, therefore, in the short term, inflation will reduce the NPL ratio and decrease the credit risk of commercial banks, while long run inflation may not have significant enough impact on NPL.

The coefficient on the share prices indices in fist difference is significant. Beck, Jakubik and Piloiu [1] have also examined that a decline of stock prices can negatively affect bank assets quality, which means the worse situation of financial market will lead to higher possibility of non-performing loans, increase the credit risk of commercial banks.

As for the yearly absolute change of unemployment rate in *t*-1 period, the NPL ratio increases following an increase in unemployment rate, which means when increasing 1% of absolute yearly change of unemployment rate in t-1 period will cause increasing of NPL ratio about 17.25%. The value of probability is 0.09, it shows unemployment growth rate has not enough significant impact on credit risk of commercial banks in selected countries.

Moreover, NPL ratio increasing with the depreciation of currency, which is also the increase of exchange rate of domestic currency to US dollar. The depreciation of domestic currency will lead to less economic activity and less credit of corporate or individual, then increase the credit risk. The p-value is less than 0.05, indicates the strongly significant impact on credit risk.

Net interest margin as one of key profitability ratio of commercial banks, although the result shows it is not significant enough for the sample, the negative sign still indicate that a high net interest margin might reflect that banks are well-capitalized to absorb credit risk, and perform high profitability.

While another profitability ratio has strongly significant impact on NPL ratio, which is ROA. It has a negative sign in the value, means the negative relationship with NPL. Higher NPL ratio in other words, worse of assets quality of bank can be caused by yearly absolute decline of profitability. The higher ROA reflects the profitability of commercial banks is strong, meanwhile, the ability to generate profit from assets can also indicate the commercial banks have effective credit risk control and regulation, therefore NPL ratio will decline.

Above discussion are based on regression model we estimated, it is the general practice of the impact of macroeconomic and bank-specific factors on credit risk management in Central European countries.

5 Conclusion

In this study, we applied econometric analysis to construct the panel regression model, through econometrical verification (autocorrelation; multicollinearity; heteroscedasticity; normality of residuals) to make sure the validity of estimated model, therefore, improve the accuracy of the model. Furthermore, we applied fixed effect to allow the intercepts to differ from country to another one. The aim of this paper is to evaluate how credit risk is influenced by macroeconomic and bank performance determinants for selected countries in Central Europe.

Based on the estimated model, we can conclude that the GDP growth rate in one-year time lag and absolute yearly change of share price indices negatively relate to the increasing of NPL ratio; Exchange rate of national currency per US dollar is positive related to the increasing of NPL ratio. The depreciation of the local currency

will lead to bad NPL ratio. Also, we have verified that bank performance indicators especially return on assets has strong influence on NPL ratio. Although unemployment rate and net interest margin has not enough significant impact on NPLs, we still can see from the result that increasing unemployment rate will lead to higher NPL ratio and the negative relationship of bank profitability and asset quality.

Furthermore, since the net interest margin is not significant and has relatively high VIF, in future research, we will examine particular cross-section that may lead to differential in credit risk management of Central European.

Acknowledgements

This work has been supported by the Czech Science Foundation (GACR) under project 19-11965S and SP2019/5, an SGS research project of VSB-TU Ostrava.

References

[1] Beck, R., Jakubik, P. and Piloiu, A. (2013) 'Non-performing loans: What matters in addition to the economic cycle?', *European Central Bank working paper series*, (1515), p. 34.

[2] Breusch, T.S. and A.R. Pagan, (1979) 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica , vol.* 47, 1287- 1294.

[3] Breusch, T., and A. Pagan. (1980) 'The Lagrange multiplier test and its application to model specification in econometrics', *Review of Economic Studies* 47: 239–253

[4] Brooks, C (2014) 'Introductory Econometrics for Finance' Cambridge University PressDougherty, C. (2011). '*Introduction to econometrics*' 4th edition Oxford: Oxford University Press.

[5] Jakubík, P. *et al.* (2008) 'Stress Testing Credit Risk: Comparison of the Czech Republic and Germany', *Financial Stability Institute FSI Award 2008 Winning Paper*.

[6] Louzis, D. P., Vouldis, A. T. and Metaxas, V. L. (2012) 'Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios', *Journal of Banking & Finance*. North-Holland, 36(4), pp. 1012–1027.

[7] Pesaran, M. H. (2004) 'General diagnostic tests for cross section dependence in panels', University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.

[8] Selma Messai, A. and Jouini, F. (2013) 'Micro and Macro Determinants of Non-performing Loans', *International Journal of Economics and Financial Issues*, 3(4), pp. 852–860.

[9] Škarica, B. (2014) 'Determinants of Non-Performing Loans in Central and Eastern European Countries', *Financial Theory and Practice, Institute of Public Finance,* vol. 38(1), pp. 37-59.

[10] Wooldridge, J. M. (2002) Econometric Analysis of Cross Section and Panel Data. *Cambridge, MA: MIT Press.*

Analysis of multiple economic internal rate of return

Stanislava Dvořáková[1], Petr Jiříček[2]

Abstract. The paper will deal with variants of finding the multiple economic internal rates of return of investment projects with a subsidy from public sources. The multiple economic internal rates of return are roots of the polynomial function that describes the investment project. We use the Descrates rule saying that the number of positive real roots of a polynomial is equal to the maximum of the number of the sign variations. The simulations will correspond to gradual changes of the subsidy amounts and the changes of cash-flows in individual investment project phases. The simulations will result in displaying functional dependencies of economic internal rates of return on multi-parametric project changes. Then the extreme values and limits of these functional dependencies will be analysed.

Keywords: multiple economic internal rate of return, ENPV function, root function, extreme values.

JEL Classification: C20, H43
AMS Classification: 65H04*2*

1 Introduction

In its mathematical basis, the efficiency evaluation of an investment project uses commercial methods – the net present value (NPV) method and the internal rate of return method (IRR). The internal rate of return method (IRR), outlined by Böhm-Bawerk [2] as roots of net present value function, was developed by Keynes [7]. Furthermore, Teichroew et al. [12] gave a complete overview of the problems of the IRR method in the past, solving the polynomial by the so-called Descartes' rule. Bussey–Eschenbach [1] define the non-conventional type of investment as an investment that contains one or more negative cash out-flows, followed by one or more positive cash inflows. In this case, in compliance with the Descartes' rule, the polynomial function used to assess investment project efficiency will have multiple roots. In case of multiple IRRs, the investment project efficiency cannot be unambiguously determined. As regards the multiple rate on return topic, professional literature deals with several approaches. Hazen [6] has solved the problem of multiple real IRRs by linking the present value of the outstanding capital expenditure of a project with the difference be-tween any IRRs and the cost of capital. The topic of using IRR and NPV criteria when multiple real-valued IRRs may arise has been summarised by Magni [8]. Pressacco et al. [10] define the so-called quasi-IRR for non-conventional projects that may not have a real–valued IRR. Pierru [11] describes real-valued and complex-valued roots of the NPV function. Dhavale–Sarkis [3] dealt with multiple roots analysis of stochastic type of internal rate of return. For the description of public project behaviour under the conditions of subsidizing from the European funds, we will use different terminology using the EU methodology (European Commission [5]): the Economic Net Present Value (ENPV) function and, for its roots, Economic Internal Rate of Return (ERR).

2 Objective and methods

The objective of this contribution is a solution of investment projects efficiency with multiple changes of cash-flow polarity, which characteristically have more than one real root of the $ENPV$ function. The simulations will result in displaying multiple roots of functional dependencies of the $ENPV$ function (multiple ERR) on negative cash-flow project changes. Then, the extreme values and limits of these functional dependencies will be determined.

The Economic Net Present Value method is expressed by the $ENPV$ function (Dvořáková–Jiříček, [4]),

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}, \tag{1}$$

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz.
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, jiricek@vspj.cz.

where

CF_t is the cash-flow of the investment project (i.e. the project inflows minus project outflows),

k is the discount rate (the required return rate of investment project),

n is the period of economic lifespan of the project (in our model case $n = 4$),

$ENPV$ is net present value of the investment project.

The constant term (CF_0) – it represents capital expenditures of the project in the model. It will acquire non-positive values, which corresponds to a negative cash-flow reflecting project investment costs or, as the case may be, various amounts of subsidy from the donor.

Coefficients (CF_1 to CF_n) – in the model represent cash-flows caused by the investment and they can generally acquire both positive and negative values

Descartes rule of signs: Under this rule, it is possible to determine how many positive roots there are for a polynomial function. The rule says that the number of positive real roots of a polynomial is equal to the maximum number of the sign variations (Novotná–Trch [9]). Projects with this variation are called non-nonconventional projects and are characterized by more than one change of project-generated cash-flows in the sequence of the CF stream (e.g.$[-,-,+,+,-]$). In our case, when considering a five-year-long project we choose three basic possibilities of non-conventional projects (along with a conventional type, see point 2). The following list enumerates the fundamental assumptions of the simulation.

1. *Stream of CFs* – the sequence of cash-flows generated by the investment project, here in the form of $CF = [CF_0, CF_1, CF_2, CF_3, CF_4]$, which corresponds to realistically predictable flows of the project.
2. In the simulation, we seek, for a unit cash-flow, for the dependence of the root curve $CF_j = \varphi_j(ERR)$ course and multiple ERR position on gradual changes of individual cash-flows due to the lifespan of the investment. The changes of cash-flow are constructed for the cases of:
   a) Project with a stream cash-flow $CF = [CF_0, -CF_1, 1, 1, 1]$.
   b) Project with a stream cash-flow $CF = [CF_0, 1, -CF_2, 1, 1]$.
   c) Project with a stream cash-flow $CF = [CF_0, 1, 1, -CF_3, 1]$.
   d) Project with a stream cash-flow $CF = [CF_0, 1, 1, 1, -CF_4]$.
3. *Economic Internal Rate of Return of the project* (ERR) – the root of the $ENPV$ function e.g. investment curve (1) of the project (Dvořáková–Jiříček [4]), i.e.

$$ENPV = CF_0 + \frac{CF_1}{1+ERR} + \frac{CF_2}{(1+ERR)^2} + \frac{CF_3}{(1+ERR)^3} + \frac{CF_4}{(1+ERR)^4} = 0 \tag{2}$$

4. In the presented study, we analyse multiple roots of the $ENPV$ function (multiple ERR) by means of modelling the defined root function.
5. *The root function* $CF_j = \varphi_j(ERR)$, $0 \le j \le n$ – the function of roots for the function of two variables of $NPV(ERR, CF_j) = 0$ (Dvořáková–Jiříček [4]), here

$$\begin{aligned}
CF_1 = \varphi_1(ERR) &= -CF_0(1+ERR) - \left(\frac{CF_2}{1+ERR} + \frac{CF_3}{(1+ERR)^2} + \frac{CF_4}{(1+ERR)^3}\right) \\
CF_2 = \varphi_2(ERR) &= -(CF_0(1+ERR)^2 + CF_1(1+ERR)) - \left(\frac{CF_3}{1+ERR} + \frac{CF_4}{(1+ERR)^2}\right) \\
CF_3 = \varphi_3(ERR) &= -\left(CF_0(1+ERR)^3 + CF_1(1+ERR)^2 + CF_2(1+ERR)\right) - \frac{CF_3}{1+ERR} \\
CF_4 = \varphi_4(ERR) &= -(CF_0(1+ERR)^4 + CF_1(1+ERR)^3 + CF_2(1+ERR)^2 + CF_3(1+ERR))
\end{aligned} \tag{3}$$

6. We use 2D projection for seeking the course of the root function and we find where we can determine the sphere of multiple ERR.

3 Results and discussions

We present a projection of a set of four root functions for selected levels of subsidy of a public project (0%, 30%, 70%, 100%) in a summary graph. In the graph, we can see the influence of the subsidy level on the course of root functions for individual scenarios and we seek multiple roots of ERR (see Fig. 1, 2, 3, 4).

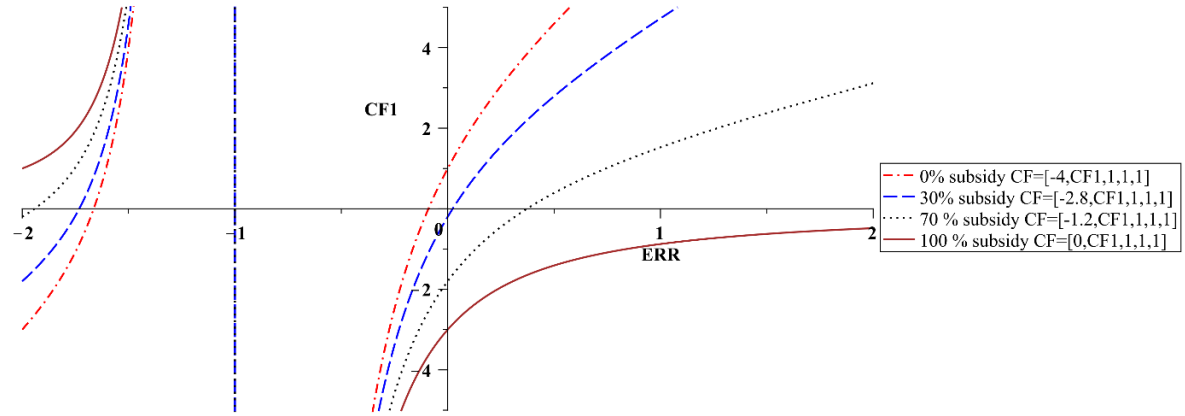


Figure 1 Quadruplet of functions $CF_1 = \varphi_1(ERR)$ for various parameters of subsidy amount

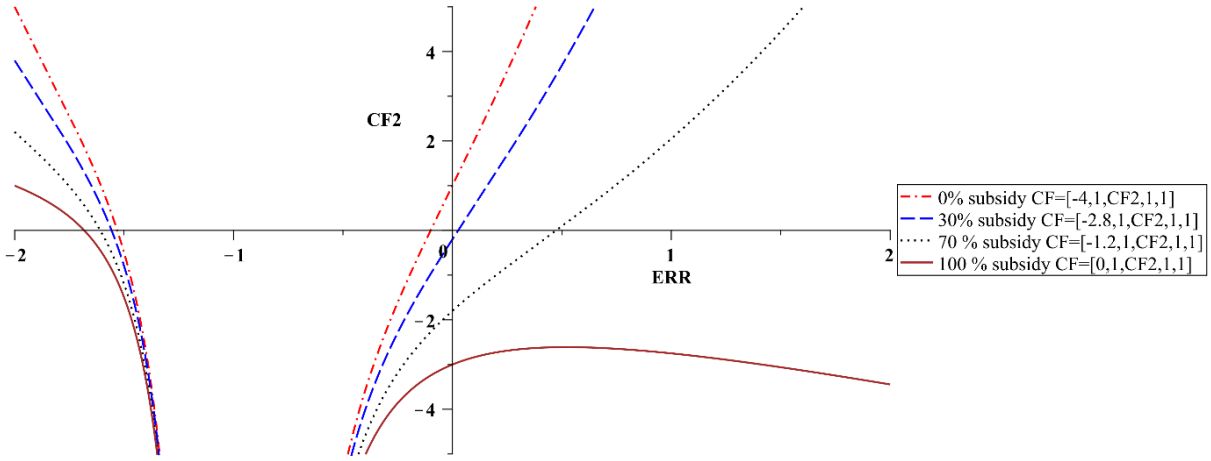


Figure 2 Quadruplet of functions $CF_2 = \varphi_2(ERR)$ for various parameters of subsidy amount

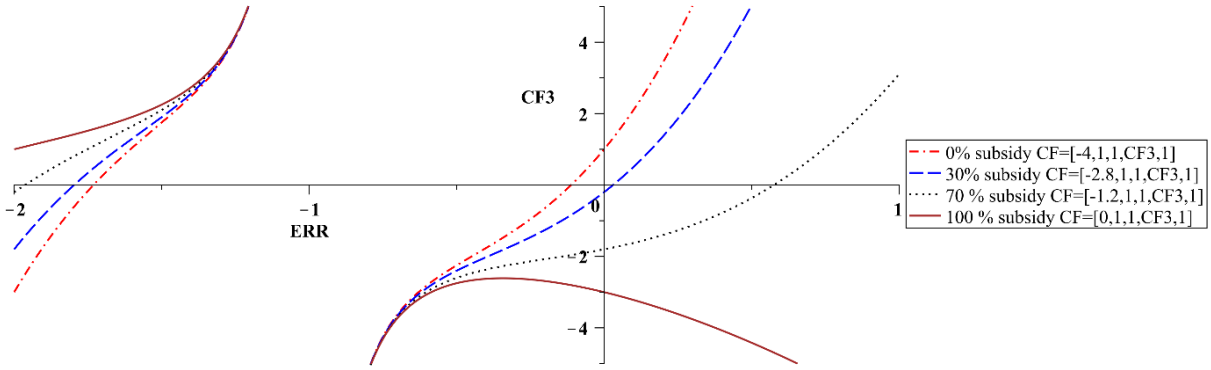


Figure 3 Quadruplet of functions $CF_3 = \varphi_3(ERR)$ for various parameters of subsidy amount

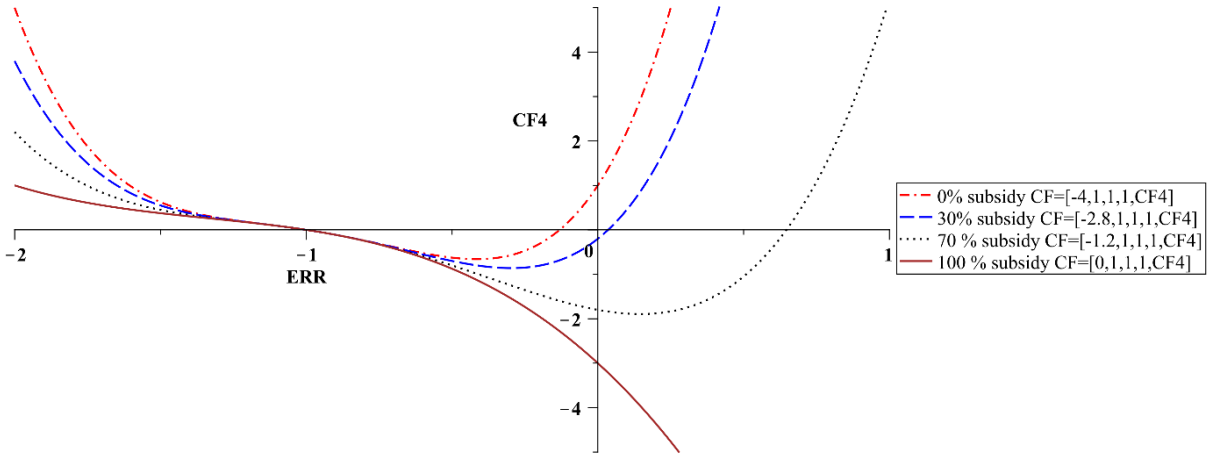


Figure 4 Quadruplet of functions $CF_4 = \varphi_4(ERR)$ for various parameters of subsidy amount

If we look at the root function $CF_2 = \varphi_2(ERR)$ for 100% subsidy in Fig. 2, we can see that, for certain values of CF_2, there are two ERR roots. In Fig. 5, there is a detail of the function. It is a non-conventional project. For a particular $CF = [0{,}1, -2{.}7{,}1{,}1]$, we draw the course of the $ENPV$ function (1), where both the roots are clearly visible where the $ENPV$ function intersects the horizontal axis. In the CF sequence, the change of the sign occurs twice; the $ENPV$ function thus has two roots of ERR ($ERR_1 = 0.235$, $ERR_2 = 0.892$) under the above mentioned Descartes' rule.

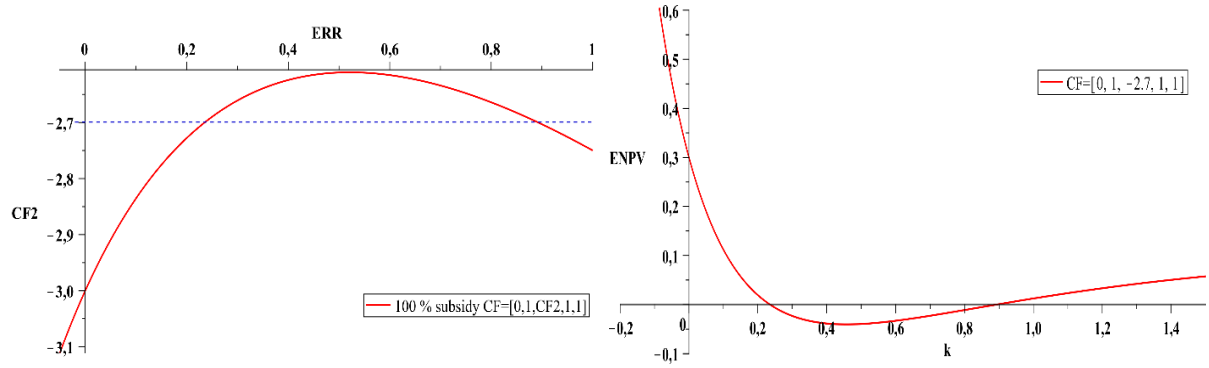


Figure 5 Detail of the root function $CF_2 = \varphi_2(ERR)$ for 70% subsidy and the course of the relevant $ENPV$ function

In Fig. 3, the situation reoccurs in the root function $CF_3 = \varphi_3(ERR)$. There are, however, two negative roots there; therefore, we do not deal with that. In the root function $CF_4 = \varphi_4(ERR)$ in Fig. 4 for 70% subsidy, we can see again that for some values of CF_4, the function $ENPV$ (1) will also have two positive ERR roots. Fig. 6 presents the situation in detail. It is a non-conventional project with two sign changes in the CF sequence again. In the particular case of $CF_4 = -1.82$, the roots are $ERR_1 = 0.017$, $ERR_2 = 0.263$.

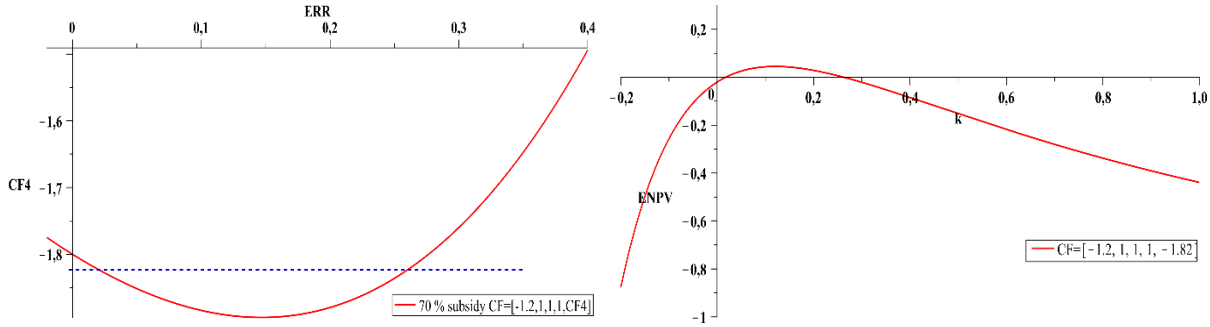


Figure 6 Detail of the root function $CF_4 = \varphi_4(ERR)$ for 70% subsidy and the course of the relevant $ENPV$ function

4 Conclusion

The ERR method is, besides the ENPV method, the principal method of evaluating the efficiency of public projects subsidized from the European funds. The disadvantage of the ERR method is the existence of multiple roots when the evaluation criterion for determining the project discount rate cannot be used unequivocally, which applies to some variants of the so called non-conventional projects. The evaluation criterion of efficiency says that the discount rate of a public investment project as the project's rate of return required by the investor or donor is limited by the amount of the Economic Internal Rate of Return ($k < ERR$). It is therefore necessary to exclude the non-conventional project variants with multiple ERR values from evaluation. In the paper, we solve the existence of the non-conventional project variant with multiple ERR roots by means of finding the extremes of the root function $CF_j = \varphi_j(ERR)$ issuing from the equation $ENPV\ (ERR, CFj) = 0$. The executed simulation for a five-year public project with various amounts of subsidy from the European funds (year 0 is the investment project stage, years 1 to 4 are the operational phase) shows that the occurrence of these non-conventional projects is limited.

Acknowledgements

This research was supported by the College of Polytechnics, Jihlava, under Grant no. 1170/26/420 Mathematical Methods in Economics.

References

[1] Bussey, L. E., Eschenbach, T. G. (1992). *The Economic Analysis of Industrial Projects*. Prentice-Hall, New Jersey

[2] Böhm-Bawerk, E. V. (1891). *The positive theory of capital*. Available at http:// mises.org/books/positivetheory.pdf.

[3] Dhavale, D. G., Sarkis, J. (2018). Stochastic internal rate of return on investment in sustainable assets generating carbon credits. *Computers & Operation Research 89*, 324–336

[4] Dvořáková, S., Jiříček, P. (2016). Multi-parametric simulation of a model of nonconventional public projects evaluated by cost-benefit analysis. In: *Proceedings 34st International Conference on Mathematical Methods in Economics* (pp. 177–182). Liberec: Technická univerzita.

[5] European Commission (2015). *Guide to Cost-Benefit Analysis of Investment Project*. Brusel: DG Regional and Urban Policy.

[6] Hazen, G. B. (2003). A new perspective on multiple internal rates on return. *The Engineering Economist 48(1)*, 31–35.

[7] Keynes, J. M. (1936). *The general theory of employment, interest and money*. Available at http://ambidextrouscivicdiscourse.com/wp-content/uploads/2010/10/The-General-Theory-of-Employment-Interest-and-Money.pdf.

[8] Magni, C. A. (2011). Average Internal Rate of Return and investment decisions: a new perspective. *The Engineering Economist 55(2)*, 150–180.

[9] Novotná J. & Trch, M. (2000). *Algebra a teoretická aritmetika. Polynomická algebra*. Praha: Univerzita Karlova.

[10] Pressacco, F., Magni, C.A., Stucchi, P. (2011). *A quasi-IRR for a project without IRR*. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1800348.

[11] Pierru, A. (2010). The simple meaning of complex rates of return. *The Engineering Economics 55(2)*, 105–117.

[12] Teichroew D, Robichek AA, Montalbano M (1965). An analysis of criteria for investment and financing decisions under certainty. *Management Science 12(3)*, 151–179.

The Evaluation and Selection of Suppliers: DEA Approach

Natalie Pelloneová[1], Eva Štichhauerová[2]

Abstract. Evaluation and selection of potential suppliers is an essential part of effective management of the current dynamic and global supply chains and is also widely discussed in the literature. Both evaluation of suppliers and selection of the best ones represent a significant and complex decision of the strategic importance that requires careful consideration of different performance criteria and attributes. One of the techniques that can be used for the evaluation and selection of suppliers is the data envelopment analysis (DEA) which uses mathematical programming to evaluate the performance of a set of homogeneous units, considering several inputs and outputs. The presented paper deals with the use of input-oriented BCC model for evaluation and subsequent selection of suppliers in the selected company. The paper is divided into three parts. The first part of the paper is devoted to the literary research focused on the use of DEA method in the evaluation and selection of suppliers. In the second part, the use of the proposed method is demonstrated on an example. The main results are summarized in the conclusion.

Keywords: Data Envelopment Analysis, DEA, supplier selection, supplier evaluation, supplier efficiency, pure technical efficiency.

JEL Classification: C14, C44, C61, L25
AMS Classification: 90C90, 90B90

1 Introduction

The evaluation and selection of suppliers is an area that is given great attention in professional literature. Suppliers form an integral part of the supply chain, and their evaluation and management require specialized negotiation skills because suppliers are not part of the organization [10]. The relationships between a company and its suppliers have always been very important. The objective of a customer's evaluation of their supplier is to provide open and objective feedback on the supplier's overall performance.

In addition to evaluating existing suppliers, one of the most important functions of the purchasing department is the selection of the best supplier from a set of potential suppliers. Most companies spend a considerable amount of their revenues on purchasing [3]. Choosing the right suppliers significantly reduces purchasing costs and increases the competitiveness of enterprises. On the other hand, choosing the wrong suppliers can cause operational and financial problems [1]. Suppliers must be carefully selected because they can have both a very positive and a very negative impact on the customer's overall performance. The evaluation and selection of a supplier are multi-criteria issues which include both qualitative and quantitative factors. Customer companies create a set of evaluating criteria to be used when comparing suppliers. The basic criteria commonly used for this purpose are: price, quality of deliveries, delivery times and level of services provided. Sometimes these evaluation criteria conflict with each other [2]. The customer then needs to find a compromise solution. Professional literature presents a fairly wide range of methods proposed for evaluating and selecting suppliers. Some of these methods include: TOPSIS, fuzzy data envelopment analysis, integration of neural networks (NN) and DEA, cross efficiency fuzzy DEA, fuzzy AHP, analytic network processes, chance constraint DEA, RDEA and DEA.

The aim of the presented paper is to prepare a procedure for evaluating the existing suppliers of a selected manufacturing enterprise and subsequently select the best from among the potential suppliers. The authors focused on the use of the data envelopment analysis method in its classic form.

2 Data Envelopment Analysis

Data envelopment analysis (DEA) is a non-parametric method based on mathematical programming that is capable of evaluating the efficiency of the decision-making units (hereinafter referred to as "DMU") of different

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz
[2] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, eva.stichhauerova@tul.cz

types. Since 1978, when it was published by Charnes, Cooper and Rhodes, a number of DEA models have been developed and a large number of applications have been described in various areas of the socio-economic behaviour of people. DMU efficiency is defined as the ratio of the weighted sum of outputs (i.e. performance) to the value of the weighted sum of inputs (i.e. resources used).

Data envelopment analysis helps the decision maker to differentiate between efficient and inefficient DMUs on the basis of quantifying the efficiency score. A score equal to one implies efficient DMUs and a score lower than one implies inefficient DMUs. Depending on the decision maker's manufacturing capabilities and the characteristics of the variables used, it is possible to choose between two basic types of DEA models: the CCR model operating with constant returns to scale and the BCC model operating with variable returns to scale [4].

2.1 Data Envelopment Analysis for Supplier Evaluation and Selection

The evaluation of existing suppliers and the selection of a new supplier form part of the strategy of improving the quality of outputs of any organization [10]. Various studies dealing with the selection and evaluation of suppliers using data envelopment analysis are briefly summarized in the next chapter according to the year in which they were published.

Kleinsorge et al. use DEA to monitor the performance of a selected supplier in 18 different time intervals. Total costs and the number of shipments are considered as the inputs, whereas the number of bills received without error, the number of shipments to arrive on time, ratings for service quality experience and confidence are encompassed as the outputs [8]. Weber uses DEA to measure the performance of a baby food manufacturer's suppliers. He states that DEA is an objective method which contrasts with the usually available subjective methods for the selection of a supplier. When making an evaluation, he uses the unit price as the input, and the reject percentage and the late delivery percentage as the outputs [15]. Narasimhan uses the CCR DEA model to evaluate 23 suppliers. Quality management practices and systems, documentation and self-audit, process/manufacturing capability, management of the firm, design and development capabilities and cost reduction capability are considered as the input variables. Quality, price, delivery, cost reduction performance and others are considered as the output variables [9]. Saen uses the CCR model and subsequently super-efficiency analysis to evaluate suppliers. Price and the number of shipments are used as the inputs for the evaluation of 12 suppliers. The number of bills received from the supplier without errors and the number of on time shipments are used as the outputs [11]. Saen builds on his paper by developing a new methodology of data envelopment analysis for selecting suppliers who have undesirable outputs and imprecise data. He evaluates 18 suppliers (DMU); the main considered input is the total costs of transportation, the output used in the study is the number of bills received from the supplier without errors [12]. Costantino et al. choose from 29 suppliers in their paper. Individual suppliers must be classified according to the MEAT criteria with two input criteria and two output criteria. The input criteria are price and implementation time, the output criteria are free maintenance post delivery and the quality of technical changes to the supply [5]. Shirouyehzad et al. illustrates the use of the DEA method using an example of 12 suppliers for the Iranian company Darakar. Price, the average number of late deliveries and the rate of rejected parts are considered as the inputs, whereas service quality is considered as the output. The DEA model identified four suppliers as efficient. The Andersen-Petersen super-efficiency model was applied to determine their order [13]. Toloo and Nalchigar evaluate 18 suppliers using the DEA method. The total costs of transportation and the supplier's reputation are considered as the inputs. The number of flawless invoices from the supplier is considered as the output [14]. Costantino et al. use the fuzzy cross-efficiency DEA model to evaluate eight potential suppliers of a medium-sized Italian enterprise which provides, installs and maintains hydraulic equipment. The aim is to select one supplier from a set of suppliers. This enterprise has never cooperated with potential suppliers, so no historical data are available. The following input data are used for the evaluation: component price, order fulfilment lead time (time elapsed from the issuance of the purchase order until its delivery), and geographical distance between the purchasing enterprise and the suppliers. The quality of supplied components and reliability of the delivery (percentage ratio of fulfilled orders to total orders) are considered as the outputs [6]. Ahmady et al. deal with the supplier selection issue in their paper and introduced the fuzzy DEA approach with double frontiers for the selection of the supplier. They show this method using an example of 18 suppliers. The model includes total transportation costs, which represent a cardinal input. Price is regarded as a fuzzy input and the supplier's reputation is a qualitative input. The number of timely delivered consignments, the number of flawlessly received invoices from the supplier and the number of delivered parts are used as the outputs [1]. Dotoli uses the stochastic cross-efficiency DEA model to select suppliers for a leading Italian forklift manufacturer. Price, lead time and distance are considered as the input criteria, whereas quality and reliability are considered as the outputs [7].

From the examples above, information can be drawn on possible input and output characteristics that can be used to evaluate the suppliers' performance.

3 Research Objective and Methodology

In this paper, the authors aim to prepare a procedure for evaluating the existing suppliers of a chosen manufacturing company and subsequently for making a selection from among the potential suppliers in a situation where it would be necessary to replace an existing supplier who has a very low rating. The procedure will be demonstrated using an illustrative example of a medium-sized manufacturing company. The company chosen in this paper will be designated with an invented name, MAXA, because the quantitative data used are fictitious at this early stage of the research. However, the authors consider this to be sufficient for the purpose of presenting the intent.

The company was founded as a limited liability company and focuses on the production of plastic parts for the automotive industry. MAXA had more than 50 suppliers in 2018. These were mostly domestic suppliers.

In the initial stage of the proposed procedure, the existing suppliers will be divided into three groups using the ABC analysis, depending on their share in the company's total turnover. The individual suppliers will be subsequently identified by numeric codes for the purpose of this paper. Group A will include key, strategically important suppliers with the highest share in the company's total turnover (altogether approximately 80% of the total turnover), Group B will include moderately important suppliers (approximately 15% of the total turnover) and Group C will include, as expected, a large number of less important suppliers. Only those existing suppliers falling under Group A will be evaluated.

For MAXA, the most important items purchased are production materials, the volume of which when expressed in monetary units, is significantly higher than other items purchased. For this reason, attention is paid below in this paper to its suppliers of production materials.

In MAXA, both the evaluation process and the supplier selection process are considered to be important factors affecting its correct operation. The evaluation concerns all key production material suppliers for the given year. The authors expect the frequency of evaluations to be carried out regularly to be once a year at the end of each calendar year and, if exceptionally necessary, to conduct another outside of this plan. It is necessary to carefully make and evaluate records of past transactions with suppliers in order to obtain the proper basis for quantifying the evaluation criteria. For MAXA, the most important input criteria are the price and quality of the production materials purchased, as the price and quality are directly reflected in the final quality and in the price of the company's products. Great emphasis is placed on the quality, timeliness and completeness of deliveries. Low-quality supplies can limit production and consequently cause complications in satisfying its customers' orders.

In the first part, the BCC-I model will be used to evaluate suppliers of production materials belonging to Group A. A total of 13 suppliers were included in Group A after making the ABC analysis. The BCC-I model works with the assumption of variable returns to scale and can be written using the formula found in (1) with a valid convexity condition (2). The BCC model provides the pure technical efficiency score (PTE).

$$\begin{aligned} \mathrm{E}_0 &= \min. \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j X_{ij} + s_i^- = \theta X_{i0}, i = 1, \dots, m \\ & \sum_{j=1}^{n} \lambda_j Y_{rj} - s_i^+ = Y_{r0}, r = 1, \dots, s \end{aligned} \tag{1}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{2}$$

$$\lambda_j, s_i^-, s_i^+ \geq 0, j = 1, \dots, n, i = 1, \dots, m, r = 1, \dots, s.$$
$$\theta \ unrestricted \ in \ sign.$$

In this model, λ_j are the weights of all DMUs, s^-_i and s^+_r are slack variables, $\varepsilon > 0$ is an infinitesimal constant defined to be smaller than any positive real number and θ is the efficiency score that expresses the reduction rate of inputs in order for this unit to reach the efficient frontier.

The first step in the evaluation process is to define the DEA model input and output variables. The following input is considered: acquisition time (i.e. the time from the moment when the need to order an item was detected until the moment the item was placed in stock). The following outputs are considered: the quality of raw materials supplied (percentage ratio of the number of quality pieces delivered to the total number of pieces

delivered), timeliness of deliveries (the number of deliveries delivered on time in any given period in relation to the total number of deliveries in that same period) and the ratio of full deliveries (the number of full deliveries in any given period in relation to the total number of deliveries in that period). The values of these criteria from 13 suppliers are listed in Table 1 in the Results section.

The BCC-I model is used to calculate the pure technical efficiency score (PTE) for each supplier in the set. A supplier rated as efficient will receive a score of 1, whereas an inefficient supplier will receive a score lower than 1. For the purpose of further processing, suppliers will be divided into four groups according to their efficiency score: 1.000 excellent supplier, 0.999–0.850 acceptable supplier, 0.849–0.700 satisfactory supplier, less than 0.699 unsatisfactory supplier. This rating is always relative with regard to the current set of suppliers.

Cooperation should be discontinued with suppliers from the unsatisfactory supplier category and they should be replaced by a new supplier who will again be selected using the BCC-I model. However, MAXA has to ask whether the supplier is replaceable. With a supplier who is not replaceable, there is no other choice but to negotiate corrective measures leading to improvement in the current situation. If the supplier is replaceable, MAXA will make a selection from among the potential suppliers. When selecting from among a set of suppliers with whom MAXA has never cooperated, the selection of the evaluation criteria is limited in terms of data availability and credibility. Data on potential suppliers can be obtained through Internet references, company price lists, catalogues and samples upon request from the supplier or through references from existing suppliers or customers who have already cooperated with the potential supplier. MAXA requires potential suppliers to submit documents such as the supplier's presentation, a list of assets, quality certificates obtained, a contingency plan, financial statements, etc.

To make the evaluation and selection of a potential supplier, criteria to be taken into account were established in cooperation with MAXA. The distance of the supplier is one of the factors affecting the final price of the product and the speed of deliveries; if an exceptional order is needed, the distance of the production material supplier could mean more downtime for the company. The authors suggest using this criterion as an input variable together with the price of material. The quality of material (that was rated by corporate experts using the seven-point Likert scale (1 = very low, 7 = very high) and maturity period for the payment of invoices, which is also relevant to MAXA, were proposed to be used as the output criteria.

4 Results

Table 1 contains the evaluation criteria for MAXA's 13 current suppliers falling under Group A. The values of the criteria in this stage of the research are only illustrative. The results of the application of the BCC-I model can also be learned from Table 1.

	Input	Outputs			PTE score	Classification
Supplier	**Acquisition time (days)**	**Quality (%)**	**Timeliness (%)**	**Completeness (%)**		
1A	19	98	95	96	1.000	excellent supplier
2A	20	97	98	96	1.000	excellent supplier
3A	13	94	96	98	1.000	excellent supplier
4A	12	98	91	95	1.000	excellent supplier
5A	21	98	96	90	1.000	excellent supplier
6A	16	96	98	96	1.000	excellent supplier
7A	20	96	95	98	1.000	excellent supplier
8A	10	95	96	95	1.000	excellent supplier
9A	14	96	97	96	1.000	excellent supplier
10A	20	90	96	98	0.650	unsatisfactory supplier
11A	10	95	97	94	1.000	excellent supplier
12A	11	95	91	95	0.909	acceptable supplier
13A	9	92	93	96	1.000	excellent supplier

Table 1 Results of the BCC-I model application

Table 1 shows that suppliers 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 11A, and 13A were rated as efficient with a PTE score equal to 1.000, which means that they are able to efficiently transform inputs into outputs. The remaining two suppliers are inefficient with a PTE score lower than 1.000. For the purpose of subsequent processing, the suppliers were further classified according to the resulting PTE score (see the last column of

Table 1). Ten excellent, one acceptable and zero satisfactory suppliers were found. Only one supplier (supplier 10A) was rated as an unsatisfactory supplier.

In the illustrative example, the situation is that the unsatisfactory supplier is replaceable, so cooperation with that supplier will be discontinued. The replacement for the missing supplier will be selected from among a set of 9 shortlisted potential suppliers. Due to the lower number of DMUs and the DEA method used, the number of evaluation criteria items had to be reduced to three. Based on consultation with a MAXA representative, the authors decided to leave out the maturity period. Table 2 shows the evaluation criteria for 9 potential suppliers together with the results of the BCC-I model application.

	Inputs		Output	PTE score	Super-efficiency score
Supplier	**Distance (km)**	**Price (CZK)**	**Quality (seven-point Likert scale)**		
P1	362	51,632	7	0.816	0.816
P2	275	43,500	7	0.968	0.968
P3	269	28,314	5	0.809	0.809
P4	235	42,108	7	1.000	1.000
P5	202	38,872	6	0.836	0.836
P6	162	37,752	6	0.861	0.861
P7	140	37,752	6	0.815	0.815
P8	133	22,904	5	1.000	1.483
P9	94	42,108	7	1.000	2.500

Table 2 Results of the BCC-I and Andersen-Petersen super-efficiency model application

It can be learned from Table 2 that suppliers P4, P8 and P9 are efficient with a PTE score equal to 1.000. The remaining six suppliers are inefficient with a PTE score lower than 1.000. For the purpose of making a more detailed evaluation of the efficient units, the Andersen and Petersen model was applied, and it can be determined on the basis of the super-efficiency scores which of these three suppliers is the best for MAXA. In the illustrative example, this was potential supplier P9 with a super-efficiency score of 2.500.

5 Conclusion

The presented paper dealt with the use of the data envelopment analysis method to help evaluate and select suppliers. The aim was to create a procedure for evaluating the existing suppliers of a chosen manufacturing company and, if necessary, subsequently also for selecting a new one from among potential suppliers. The invented name MAXA was intentionally used for the company because illustrative data were used in the initial stage of the research carried out. The authors of the paper plan to verify the established procedure in practice during the next stage of research.

The proposed procedure was divided into three stages. In the first stage, MAXA's suppliers were classified using the ABC analysis tool and suppliers from Group A, who account for approximately 80% of the value of the purchased stock of production materials, were identified. In the illustrative example in this paper, this comprised a set of 13 suppliers. In the second stage, these suppliers were evaluated through the BCC-I model using one input and three outputs. The result of the evaluation was the classification of suppliers from Group A according to the achieved efficiency score.

The third stage of the proposed procedure depended on whether any of the existing suppliers had been classified as unsatisfactory. The BCC-I model was applied again, this time with the aim of evaluating potential suppliers and selecting one that would be an adequate replacement for the supplier with whom cooperation was discontinued. The inputs and outputs used depended on what information could be obtained about the suppliers with whom MAXA has not yet cooperated. Compared to other multi-criteria methods of evaluating variants, the DEA method is also limited by the fact that the maximum number of criteria that can be used depends on the size of the set of compared variants. The smaller the set of evaluated suppliers, the fewer criteria that can be included in the evaluation. Since nine potential suppliers were identified in the illustrative example in this paper, only three evaluation criteria could be included in the model: two inputs and one output. The result of applying the DEA method was the obtaining of the efficiency score of all entities. A suitable supplier from the several suppliers identified as efficient was then selected based on the highest super-efficiency score.

When selecting suppliers, a company also needs take into account other aspects that cannot be easily included in the evaluation. In practice, for example, a situation could occur where a potential supplier is evaluated as unsatisfactory and MAXA would be able to identify a more suitable partner for the given order, but

an important customer directly requests that parts be specifically purchased from this supplier selected by the customer. Through this example, MAXA shows that it respects the wishes of its customers and offers them a certain competitive advantage by meeting their requirements. This is important because a satisfied customer returns and concludes other contracts.

Acknowledgements

Supported by the TUL – Technical University of Liberec, Faculty of Economics under SGS grant No. 21294 "The use of quantitative methods in evaluating the performance of economic entities".

References

[1] Ahmady, N., Azadi, M., Sadeghi, S. A. H. & Saen, R. F. (2013). A novel fuzzy data envelopment analysis model with double frontiers for supplier selection. *International Journal of Logistics Research and Applications*, 16(2), 87–98.

[2] Azadeh, A. & Alem, S. M. (2010). A flexible deterministic, stochastic and fuzzy Data Envelopment Analysis approach for supply chain risk and vendor selection problem: Simulation analysis. *Expert Systems with Applications*, 37, 7438–7448.

[3] Çebi, F. & Bayraktar, D. (2003). An integrated approach for supplier selection. *Logistics Information Management*, 16(6), 395–400.

[4] Charnes, A., Cooper, W. W., Lewin, A. Y. & Seiford, L. M. (1994). Data Envelopment Analysis: Theory, Methodology, and Applications. Norwell: Kluwer Academic Publishers.

[5] Costantino, N., et al. (2011). Supplier selection in the public procurement sector via a Data Envelopment Analysis approach. *Proc. 19th IEEE Mediterranean Conf. on Control and Automation (MED 2011)*, Corfu, Greece, June 23-25, 2011.

[6] Costantino, N., et al. (2012). A cross efficiency fuzzy Data Envelopment Analysis technique for supplier evaluation under uncertainty. *Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, Krakow, Poland, September 17-21, 2012.

[7] Dotoli, M., Epicoco, N., Falagario, M. & Sciancalepore, F. (2015). A Stochastic Cross-efficiency Data Envelopment Analysis Approach for Supplier Selection under Uncertainty. *International Transactions in Operational Research*, 23, 725–748.

[8] Kleinsorge, I. K., Schary, P. B. & Tanner, R. D. (1992). Data envelopment analysis for monitoring customer-supplier relationships. *Journal of Accounting and Public Policy*, 11, 357–372.

[9] Narasimhan, R., Talluri, S. & Mendez, D. (2001). Supplier Evaluation and Rationalization via Data Envelopment Analysis: An Empirical Examination. *The Journal of Supply Chain Management*, 37(3), 28–37.

[10] Ramanathan, R. (2007). Supplier selection problem: integrating DEA with the approaches of total cost of ownership and AHP. *Supply Chain Management: An International Journal*, 12(4), 258–261.

[11] Saen, R. F. (2008). Using super-efficiency analysis for ranking suppliers in the presence of volume discount offers. *International Journal of Physical Distribution & Logistics Management*, 38(8), 637–651.

[12] Saen, R. F. (2010). Developing a new data envelopment analysis methodology for supplier selection in the presence of both undesirable outputs and imprecise data. *The International Journal of Advanced Manufacturing Technology*,51, 1243–1250.

[13] Shirouyehzad, H., et al. (2011). Efficiency and Ranking Measurement of Vendors by Data Envelopment Analysis. *Canadian Center of Science and Education*, 4(2), 137–146.

[14] Toloo, M. & Nalchigar, S. (2011). A new DEA method for supplier selection in presence of both cardinal and ordinal data. *Expert Systems with Applications*, 38, 14726–14731.

[15] Weber, C.A. (1996). A data envelopment analysis approach to measuring vendor performance. *Supply Chain Management*, 1(1), 28–39.

Fuzzy interval Monge matrices with respect to robustness

M. Molnárová[1]

Abstract. Periodic properties of interval Monge matrices over fuzzy algebra are studied. An interval Monge matrix over fuzzy algebra is a set of Monge matrices given by a lower bound Monge matrix and an upper bound Monge matrix. We can consider possible robustness, if there is at least one robust matrix A in the set or we can consider universal robustness, if all matrices A in the set are robust. In this paper we prove sufficient conditions for possible robustness of interval Monge matrices in case of trivial lower bound matrix. Further, we introduce necessary conditions for universal robustness of an interval Monge matrix in general case. Moreover, sufficient conditions for universal robustness of interval Monge matrices in case of non-trivial with respect to interval matrix lower bound matrix were proved.

Keywords: (max, min) algebra, robustness, interval matrix, Monge matrix

JEL classification: C02
AMS classification: 08A72, 90B35, 90C47

1 Introduction

The max-min algebra is an extremal algebra (at least one operation creates no new elements) used to model applications of discrete dynamic systems. The max-min algebra called also fuzzy algebra is used in diverse areas (graph theory, knowledge engineering, managing traffic or production) where the considered systems or devices can be represented by a matrix. Properties of fuzzy matrices were described in [9]. The Monge matrices and their applications were studied in [1], [4]. Robustness of Monge fuzzy matrices was presented in [6]. Robustness of interval fuzzy matrices was studied in [8] and [7].

2 Background of the problem

The fuzzy algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus$, $\otimes$. The least element in B will be denoted by O, the greatest one by I. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$, the least common multiple of the set S is denoted by $\operatorname{lcm} S$. For a given natural $n \in \mathbb{N}$, we use the notation $N = \{1, 2, \ldots, n\}$. For any $m, n \in \mathbb{N}$, $B(m, n)$ denotes the set of all matrices of type $m \times n$ and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. The matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes$) as matrix operations over any field. The rth power of a matrix $A \in B(n, n)$ is denoted by A^r, with elements a_{ij}^r. For $A, C \in B(n, n)$ we write $A \leq C$ if $a_{ij} \leq c_{ij}$ holds for all $i, j \in N$. For a given matrix $A \in B(n, n)$ the symbol $G(A) = (N, E)$ stands for the complete, edge–weighted digraph associated with A, i.e. the vertex set of $G(A)$ is N, and the capacity of any edge $(i, j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A, h)$ is the digraph $G = (N, E')$ with the vertex set N and the edge set $E' = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$.

The following lemma describes the relation between matrices and corresponding threshold digraphs.

Lemma 1. [8] 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 $G(A, h_2) \subseteq G(A, h_1)$.

[1]Technical University of Košice, Department of Mathematics and Theoretical Informatics, Němcovej 32, 04200 Košice, Slovakia, e-mail: Monika.Molnarova@tuke.sk

A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. The *period* of $\mathcal{K}$ is defined as $\operatorname{per}\mathcal{K} = \gcd\{\,\ell(c);\ c \text{ is a cycle in } \mathcal{K} \text{ with length } \ell(c) > 0\,\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per}\mathcal{K} = 1$. $\mathrm{SCC}^\star(G)$ is the set of all non-trivial strongly connected components of G.

Let $A \in B(n,n)$ and $x \in B(n)$. The sequence $O(A,x) = \{x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots\}$ is the orbit of $x = x^{(0)}$ generated by A, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.
For a given matrix $A \in B(n,n)$, the number $\lambda \in B$ and the n–tuple $x \in B(n)$ are the so-called *eigenvalue* of A and *eigenvector* of A, respectively, if they satisfy the equation $A \otimes x = \lambda \otimes x$. We define the corresponding *eigenspace* $V(A,\lambda)$ as the set $V(A,\lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.
Let $\lambda \in B$. A matrix $A \in B(n,n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \geq R$. The smallest natural number p with above property is called the period of A, denoted by $\operatorname{per}(A,\lambda)$. In case $\lambda = I$ we denote $\operatorname{per}(A,I)$ by abbreviation $\operatorname{per} A$.

Definition 1. Let $A = (a_{ij}) \in B(n,n)$, $\lambda \in B$. Let $T(A,\lambda) = \{x \in B(n);\ O(A,x) \cap V(A,\lambda) \neq \emptyset\}$. A is called *λ–robust* if $T(A,\lambda) = B(n)$. A λ–robust matrix with $\lambda = I$ is called *a robust matrix*.

Lemma 2. [9] Let $A = (a_{ij}) \in B(n,n)$. Then A is robust if and only if $\operatorname{per} A = 1$.

3 Robustness of Monge matrices

Definition 2. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is a convex Monge matrix if and only if

$$a_{ij} \otimes a_{kl} \leq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l.$$

Equivalent conditions for robustness of Monge matrices in general case of max-min algebra was proved in [6]. The existence of cycles of odd length, especially loops, in the corresponding threshold digraphs is crucial for investigation of robustness of matrices. We recall the corresponding lemmas. Obviously it is enaugh to consider only threshold digraphs $G(A,h)$ for $h \in H = \{a_{ij};\ i,j \in N\}$.

Lemma 3. [4] Let $A \in B(n,n)$ be a Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^\star(G(A,h))$. Let $c = (i_1, i_2, \dots, i_k, i_1)$ be a cycle of length $\ell(c) \geq 3$ in $\mathcal{K}$. Then there are arcs (i_j, i_{j+1}) and (i_l, i_{l+1}) in c such that

$$i_j < i_l \quad \text{and} \quad i_{j+1} < i_{l+1}.$$

Remark 1. By Lemma 3 each cycle of length at least three can be replaced by pair of shorter cycles. Repeated applying of the procedure for an arbitrary cycle of length at least three results in a set of cycles with length at most two. Hence we can restrict ourselves in our considerations to cycles of length one and two.

Lemma 4. [6] Let $A \in B(n,n)$ be a Monge matrix. Let $h \in H$. Let for $i, k \in N$ be the loops (i,i) and (k,k) in the digraph $G(A,h)$. Then the nodes i and k are in the same non-trivial strongly connected component $\mathcal{K}$ of $G(A,h)$.

Theorem 1. [6] *Let $A \in B(n,n)$ be a Monge matrix. Then A is robust if and only if for each $h \in H$ the digraph $G(A,h)$ contains at most one non-trivial strongly connected component and this has a loop.*

4 Robustness of interval Monge matrices

Similarly to [2], [5], we define an interval matrix $\mathbf{A}$.

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. An interval matrix $\mathbf{A}$ is called

- *possibly robust* if there exists a matrix $A \in \mathbf{A}$ such that A is robust,
- *universally robust* if each matrix $A \in \mathbf{A}$ is robust.

Definition 5. An interval matrix $\mathbf{A}^M$ for $\mathbf{A} = [\underline{A}, \overline{A}]$ is called interval Monge, if $\underline{A}$, $\overline{A} \in B(n,n)$ are Monge matrices and $\mathbf{A}^M = \{A \in \mathbf{A};\ A \text{ is Monge}\}$.

We have proved a necessary condition for an interval Monge matrix to be possibly robust in [7]. In case of interval matrices it is enaugh to consider threshold digraphs $G(A,h)$ for $h \in H = \{\overline{a}_{ij};\ i,j \in N\} \cup \{\underline{a}_{ij};\ i,j \in N\}$.

Theorem 2. [7] *If an interval Monge matrix $\mathbf{A}^M$ is possibly robust then for each $h \in H$ holds: if the digraph $G(\overline{A},h)$ contains a strongly connected component $\mathcal{K}^\star$ with a loop then this is unique and all non-trivial strongly connected components of the digraph $G(\underline{A},h)$ are subdigraphs of $\mathcal{K}^\star$.*

Definition 6. We say, that a matrix A is a trivial matrix, if the threshold digraph $G(A,h)$, for each $h \in \{a_{ij};\ i,j \in N\} - \{0\}$, contains no cycle of positive length.

Theorem 3. *Let $\mathbf{A} = [\underline{A}, \overline{A}]$. Let the matrix $\underline{A}$ be trivial. Then the interval Monge matrix $\mathbf{A}^M$ is possibly robust.*

Proof. The assertion of the theorem follows by Definition 6 and Theorem 1. □

Theorem 4. *Let $\mathbf{A}^M$ be an interval Monge matrix for $\mathbf{A} = [\underline{A}, \overline{A}]$ with $\underline{A}, \overline{A} \in B(n,n)$. Let $h^\star$ be a threshold such that both conditions hold*

(i) $G(\overline{A}, h^\star)$ contains cycle $(1,n,1)$,

(ii) $G(\underline{A}, h^\star)$ contains no cycle of positive length.

Then $\mathbf{A}^M$ is not universally robust.

Proof. Let $h^\star$ be a threshold for which the conditions (i) and (ii) hold. We shall show that there is a non-robust matrix $A^\star \in \mathbf{A}^M$. We define $A^\star$ as follows

$$a^\star_{ij} = \begin{cases} \overline{a}_{ij} & \text{if } i = 1 \text{ and } j = n, \\ \overline{a}_{ij} & \text{if } i = n \text{ and } j = 1, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases} \tag{1}$$

First, we shall verify that $A^\star$ is a Monge matrix. Using the Monge property of the matrix $\underline{A}$ we can show that for $1 \le j < n$ and $1 < k \le n$ holds $a^\star_{1j} \otimes a^\star_{kn} \le a^\star_{1n} \otimes a^\star_{kj}$:

$$a^\star_{1j} \otimes a^\star_{kn} = \underline{a}_{1j} \otimes \underline{a}_{kn} \le \underline{a}_{1n} \otimes \underline{a}_{kj} \le \overline{a}_{1n} \otimes \underline{a}_{kj} = a^\star_{1n} \otimes a^\star_{kj}.$$

Similarly we can show that for $1 \le i < n$ and $1 < l \le n$ holds $a^\star_{i1} \otimes a^\star_{nl} \le a^\star_{il} \otimes a^\star_{n1}$:

$$a^\star_{i1} \otimes a^\star_{nl} = \underline{a}_{i1} \otimes \underline{a}_{nl} \le \underline{a}_{il} \otimes \underline{a}_{n1} \le \overline{a}_{il} \otimes \underline{a}_{n1} = a^\star_{il} \otimes a^\star_{n1}.$$

In rest of cases, i.e. neigher $a^\star_{1n}$ nor $a^\star_{n1}$ is involved, the Monge property of $\underline{A}$ guaranties the inequality due to Definition 2. Hence, $A^\star$ is a Monge matrix.

Second, we shall show, that $A^\star$ is a non-robust matrix. The threshold digraph $G(A^\star, h^\star)$ contains exactly one non-trivial strongly connected component, namely the cycle $(1,n,1)$ of length two. Due to Theorem 1 $A^\star$ is not robust.

Hence, $\mathbf{A}^M$ is not universally robust. □

Example 1. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}, \overline{A} \in B(5,5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 2 & 2 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \end{pmatrix}. \qquad A^\star = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The Monge property of both matrices $\underline{A}$ and $\overline{A}$ guarantees that the corresponding interval matrix $\mathbf{A}^M$ is an interval Monge matrix. There is $h^\star = 2$ such that $G(\overline{A}, h^\star)$ contains cycle $(1,n,1) = (1.5.1)$, while $G(\underline{A}, h^\star)$ contains no cycle of positive length. Indeed there is a matrix $A^\star \in \mathbf{A}^M$ defined by (1) which is not robust (see Figure 1).

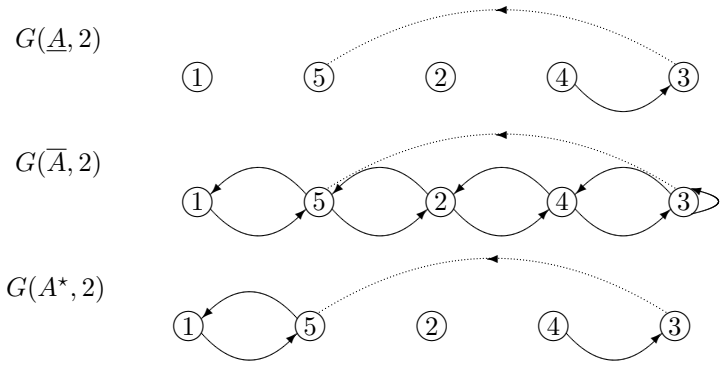


Figure 1: Necessary condition for universal robustness

Definition 7. We say, that a matrix $A \in \mathbf{A}$ is a non-trivial matrix with respect to $\mathbf{A}$, if the threshold digraph $G(A, h^\star)$, for $h^\star = \max\{\overline{a}_{ij};\, i, j \in N\}$, contains a cycle of positive length. Otherwise A is trivial with respect to $\mathbf{A}$.

Theorem 5. *Let $\mathbf{A} = [\underline{A}, \overline{A}]$. Let the matrix $\underline{A}$ be non-trivial with respect to $\mathbf{A}$. Then the interval Monge matrix $\mathbf{A}^M$ is universally robust, if for each $h \in H$ the following conditions hold*

(i) $G(\overline{A}, h)$ contains exactly one non-trivial strongly connected component $\overline{\mathcal{K}}_h = (N_{\overline{\mathcal{K}}_h}, E_{\overline{\mathcal{K}}_h})$ and this component has a loop,

(ii) $G(\underline{A}, h)$ contains exactly one non-trivial strongly connected component $\underline{\mathcal{K}}_h = (N_{\underline{\mathcal{K}}_h}, E_{\underline{\mathcal{K}}_h})$ and this component has a loop,

(iii) Let $t = \min\{N_{\underline{\mathcal{K}}_h}\}$ and $u = \max\{N_{\underline{\mathcal{K}}_h}\}$. For each cycle (m, n, m) in $\overline{\mathcal{K}}_h$, with $m < t$, $n > u$, there exists at least one of the indices b, c and at least one of the indices d, e, for b, $d \le m$ and c, $e \ge n$ such that both conditions hold

- *$\underline{a}_{bu} \ge h$ or $\underline{a}_{ct} \ge h$,*
- *$\underline{a}_{te} \ge h$ or $\underline{a}_{ud} \ge h$.*

Proof. Let the conditions (i) and (ii) hold for each $h \in H$. Hence by Theorem 1 are the matrices $\underline{A}$ and $\overline{A}$ robust. We shall show that each matrix $A \in \mathbf{A}^M$ such that $\underline{A} < A < \overline{A}$ is robust. Let the condition (iii) holds for arbitrary but fixed $h \in H$. We shall consider three cases.

Case 1. Let the non-trivial strongly connected component $\overline{\mathcal{K}}_h$ be generated by one node. Consequently $\underline{\mathcal{K}}_h = \overline{\mathcal{K}}_h$. Moreover, $G(A, h)$ contains exactly one non-trivial strongly connected component $\mathcal{K}_h = \overline{\mathcal{K}}_h$ and this has a loop. Hence A is robust.

Case 2. Let the non-trivial strongly connected component $\overline{\mathcal{K}}_h$ be generated by two nodes. Then $\underline{\mathcal{K}}_h \subseteq \overline{\mathcal{K}}_h$ and by condition (ii) $\underline{\mathcal{K}}_h$ contains at least one node and this has a loop. Due to Lemma 4 and Lemma 1 the digraph $G(A, h)$ contains exactly one non-trivial strongly connected component $\mathcal{K}_h$ such that $\underline{\mathcal{K}}_h \subseteq \mathcal{K}_h \subseteq \overline{\mathcal{K}}_h$ and this has a loop. Hence A is robust.

Case 3. Let the non-trivial strongly connected component $\overline{\mathcal{K}}_h$ be generated by at least three nodes. Let $t = \min\{N_{\underline{\mathcal{K}}_h}\}$ and $u = \max\{N_{\underline{\mathcal{K}}_h}\}$. Let for each cycle (m, n, m) in $\overline{\mathcal{K}}_h$, with $m < t$, $n > u$, there exist at least one of the indices b, c and at least one of the indices d, e, for b, $d \le m$ and c, $e \ge n$ such that both conditions in (iii) hold.
The digraph $G(A, h)$ contains a non-trivial strongly connected component $\mathcal{K}_h = (N_{\mathcal{K}_h}, E_{\mathcal{K}_h})$ such that $\underline{\mathcal{K}}_h \subseteq \mathcal{K}_h$. We will prove that if the considered cycle (m, n, m) with above properties is in $G(A, h)$ then m, $n \in N_{\mathcal{K}_h}$. Thus by Lemma 4 and Remark 1 $G(A, h)$ contains exactly one non-trivial strongly connected component $\mathcal{K}_h$ such that $\underline{\mathcal{K}}_h \subseteq \mathcal{K}_h \subseteq \overline{\mathcal{K}}_h$ and this has a loop.
In cases when equality appears for at least one element from both pairs b, c and d, e, respectively, in inequalities b, $d \le m$ or c, $e \ge n$, respectively, is the cycle (m, n, m) in the non-trivial strongly connected component $\underline{\mathcal{K}}_h$. This is a contradiction to the assumption $m < t$, $n > u$. Excluding those possibilities we shall consider four cases.
First. Let there exist $b \le m$ and $e \ge n$ such that $\underline{a}_{bu} \ge h$ and $\underline{a}_{te} \ge h$, i.e. there exist arcs (b, u) and (t, e) in $G(A, h)$.

- Let $b = m$ and $e > n$. Hence the arc (m, u) is by assumption in $G(A, h)$. Moreover, by the monge property of A and $\underline{A} \leq A$ follows

$$h \leq a_{mn} \otimes a_{te} \leq a_{me} \otimes a_{tn}.$$

  Thus $a_{tn} \geq h$, i.e. the arc (t,n) is in $G(A, h)$ and consequently $m,\, n \in N_{\mathcal{K}_h}$.

- Let $b < m$ and $e = n$. Hence the arc (t, n) is by assumption in $G(A, h)$. Moreover, by the monge property of A and $\underline{A} \leq A$ follows

$$h \leq a_{bu} \otimes a_{mn} \leq a_{bn} \otimes a_{mu}.$$

  Thus $a_{mu} \geq h$, i.e. the arc (m,u) is in $G(A, h)$ and consequently $m,\, n \in N_{\mathcal{K}_h}$.

- Let $b < m$ and $e > n$. By the monge property of A and $\underline{A} \leq A$ follows

$$h \leq a_{bu} \otimes a_{mn} \leq a_{bn} \otimes a_{mu},$$

$$h \leq a_{mn} \otimes a_{te} \leq a_{me} \otimes a_{tn},$$

  Thus $a_{mu} \geq h$, i.e. the arc (m,u) is in $G(A, h)$ and $a_{tn} \geq h$, i.e. the arc (t,n) is in $G(A, h)$. Consequently $m,\, n \in N_{\mathcal{K}_h}$.

Second. Let there exist $b \leq m$ and $d \leq m$ such that $\underline{a}_{bu} \geq h$ and $\underline{a}_{ud} \geq h$, i.e. there exist arcs (b, u) and (u, d) in $G(A, h)$. By analogy to first case we can prove the existence of arcs (u, m) and (m, u).
Third. Let there exist $c \geq n$ and $e \geq n$ such that $\underline{a}_{ct} \geq h$ and $\underline{a}_{te} \geq h$, i.e. there exist arcs (c, t) and (t, e) in $G(A, h)$. By analogy to first case we can prove the existence of arcs (n, t) and (t, n).
Fourth. Let there exist $c \geq n$ and $d \leq m$ such that $\underline{a}_{ct} \geq h$ and $\underline{a}_{ud} \geq h$, i.e. there exist arcs (c, t) and (u, d) in $G(A, h)$. By analogy to first case we can prove the existence of arcs (u, m) and (n, t). □

Example 2. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A},\, \overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 2 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},\ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 2 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

The Monge property of both matrices $\underline{A}$ and $\overline{A}$ guarantees that the corresponding interval matrix $\mathbf{A}^M$ is an interval Monge matrix. It is enaugh to consider threshold digraphs for $h \in H = \{1, 2\}$. Both $G(\underline{A}, h)$ and $G(\overline{A}, h)$ contain eactly one non-trivial strongly connected component and this has a loop for each $h \in H$. Thus both matrices $\underline{A}$ and $\overline{A}$ are robust.
For $h = 1$ the only non-trivial strongly connected component in $G(\underline{A}, 1)$ is $\underline{\mathcal{K}}_1 = (N_{\underline{\mathcal{K}}_1}, E_{\underline{\mathcal{K}}_1})$ and this is generated by the node set $N_{\underline{\mathcal{K}}_1} = \{3, 4\}$ (see Figure 2). Hence $t = \min\{N_{\underline{\mathcal{K}}_1}\} = 3$ and $u = \max\{N_{\underline{\mathcal{K}}_1}\} = 4$. There is only one cycle $(m, n, m) = (2, 5, 2)$ in the only non-trivial strongly connected component $\overline{\mathcal{K}}_1 = (N_{\overline{\mathcal{K}}_1}, E_{\overline{\mathcal{K}}_1})$ generated by the node set $N_{\overline{\mathcal{K}}_1} = \{2, 3, 4, 5\}$ in $G(\overline{A}, 1)$ such that $m < t$ and $n > u$. There exist indices $b = 1 \leq m$ and $d = 2 \leq m$ such that $\underline{a}_{bu} = \underline{a}_{14} \geq h = 1$ and $\underline{a}_{ud} = \underline{a}_{42} \geq h = 1$. The Monge property of an arbitrary matrix $A \in \mathbf{A}^M$ such that the digraph $G(A, 1)$ contains the cycle $(2, 5, 2)$ guaranties that the digraph contains the arc $(2, 4)$

$$1 = h \leq a_{14} \otimes a_{25} \leq a_{15} \otimes a_{24}.$$

Since $a_{42} \geq \underline{a}_{42} \geq 1$ the digraph $G(A, 1)$ contains the arc $(4, 2)$ as well. Hence there is exactly one non-trivial strongly connected component in $G(A, 1)$ and this has a loop on node 3.
For $h = 2$ the only non-trivial strongly connected component in $G(\underline{A}, 2)$ is $\underline{\mathcal{K}}_2 = (N_{\underline{\mathcal{K}}_2}, E_{\underline{\mathcal{K}}_2})$ and this is generated by the node set $N_{\underline{\mathcal{K}}_2} = \{3\}$ (see Figure 2). Hence $t = \min\{N_{\underline{\mathcal{K}}_2}\} = 3$ and $u = \max\{N_{\underline{\mathcal{K}}_2}\} = 3$. There is only one cycle $(m, n, m) = (2, 4, 2)$ in the only non-trivial strongly connected component $\overline{\mathcal{K}}_2 = (N_{\overline{\mathcal{K}}_2}, E_{\overline{\mathcal{K}}_2})$ generated by the node set $N_{\overline{\mathcal{K}}_2} = \{2, 3, 4\}$ in $G(\overline{A}, 2)$ such that $m < t$ and $n > u$. There exist indices $c = 4 \geq n$ and $e = 5 \geq n$ such that $\underline{a}_{ct} = \underline{a}_{43} \geq h = 2$ and $\underline{a}_{te} = \underline{a}_{35} \geq h = 2$. The Monge property of an arbitrary matrix $A \in \mathbf{A}^M$ such that the digraph $G(A, 2)$ contains the cycle $(2, 4, 2)$ guaranties that the digraph contains the arc $(3, 4)$

$$2 = h \leq a_{24} \otimes a_{35} \leq a_{25} \otimes a_{34}.$$

Since $a_{43} \geq \underline{a}_{43} \geq 2$ the digraph $G(A, 2)$ contains the arc $(4, 3)$ as well. Hence there is exactly one non-trivial strongly connected component in $G(A, 2)$ and this has a loop on node 3.
Consequently A is robust.

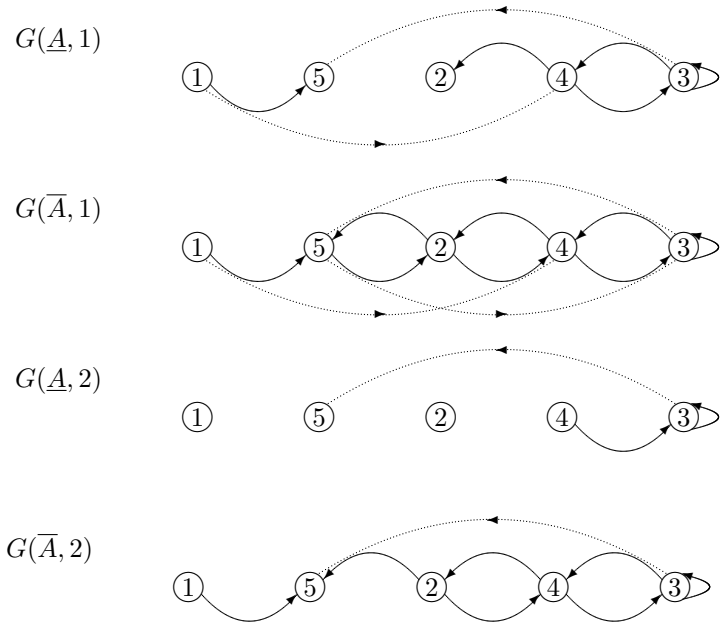


Figure 2: Sufficient condition for universal robustness

5 Conclusion

The aim of this paper is to present results concerning possible and universal robustness of interval Monge matrices over max-min algebra.

The theory of matrix robustness can reflect the following economic background. The fuzzy matrix $A = (a_{ij}) \in B(n, n)$ can represent the preferences of a customer due to different properties of a commodity on the market. The level of each property i is given by the value x_i, influenced by all properties x_j. The infuence is represented by a factor a_{ij}. The vector x at time $r + 1$ can be computed by equation $x(r + 1) = A \otimes x(r)$. The resulting eigenvector $y = A^k \otimes x$ is the maximum stable vector, i.e. the final vector of preferences. Replacing the fixed matrix A by an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ allows to find at least one robust matrix within minimum $\underline{A}$ and maximum $\overline{A}$ restrictions.

References

[1] Burkard, R. E., Klinz, B. and Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.

[2] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J. and Zimmermann, K.: Linear Optimization Problems with Inexact Data. Springer–Verlag, Berlin 2006.

[3] Gavalec, M.: *Periodicity in extremal algebras*. GAUDEAMUS, Hradec Králové, 2004.

[4] Gavalec, M. and Plavka, J.: An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max algebra, *DAM*, Volume **127** (2003), 651–656.

[5] Gavalec, M. and Zimmermann, K.: Classification of solutions to systems of two-sided equations with interval coefficients, *Inter. J. of Pure and Applied Math.* Volume **45** (2008), 533–542.

[6] Molnárová, M.: Robustness of Monge matrices in fuzzy algebra, In: *Proceedings of* 32^{nd} *Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.

[7] Molnárová, M.: Possible and universal robustness of Monge fuzzy matrices, In: *Proceedings of* 34^{th} *Int. Conference Mathematical Methods in Economics 2016*, Liberec, 2016, 582–587.

[8] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[9] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue 5 (2011), 381–388.

Problem of trade-offs between portfolio's mean, variance and skewness as a goal programming model

Renata Dudzińska-Baryła[1], Donata Kopańska-Bródka[2], Ewa Michalska[3]

Abstract. The assumption regarding a symmetrical distribution of the rate of return in the Markowitz portfolio selection model is unrealistic and has no empirical evidence. Thus the higher order moments of the portfolio's return should be taken into account. Investors prefer high values for odd moments and low ones for even moments. In this paper, we consider the goal programming model in which the weight of each goal represents the investor preference towards mean, variance and skewness of the portfolio's return. Our aim is to analyze the portfolios from the trade-off lines (frontiers) determined for various weights. The empirical study is based on quotations of stocks listed on the Warsaw Stock Exchange.

Keywords: goal programming, trade-off line, skewness, portfolio, mean-variance-skewness portfolio, investor's preferences.

JEL Classification: G11
AMS Classification: 91G10

1 Introduction

The Portfolio Selection Model introduced by Markowitz is based on two moments of the distribution of rates of return: expected value and variance. This model offers the most popular and general trade-off between risk and expected return when the return rates of stocks are normally distributed. Unfortunately, there has been no empirical evidence that rates of return are symmetrically distributed [8, 17], hence the higher order moments of the rates of return cannot be ignored. Analysts basing on the Markowitz portfolio selection model have focused on the expected rate of return and variance and have stated that an investor should always choose an efficient portfolio. The main difficulty with the approaches where the higher moments are considered is that the dimensionality of the portfolio selection problem increases, and the geometric projection of the efficient portfolios frontier is not trivial [2, 10].

The portfolio selection models which incorporate the higher moments are considered as multi-objective optimization problems. In general, investors prefer high values for odd moments and low ones for even moments. Increasing the positive value of an odd moment can be interpreted as decreasing extreme values on the side of losses and increasing them on the gains' side. Scott and Horvath [19] have shown that if the distribution of random rates of return is asymmetric or the investor's utility function is a function of a higher degree than the quadratic function, the assessment of the investment should be based on at least the third and fourth order moments.

The first publications in which portfolio selection models incorporating the third moment were proposed, appeared in the 1970s [20]. Maximizing skewness measured by the third moment expresses the basic preferences of the decision maker and leads to an increase in the chances of achieving above-average rates of return. Currently, different approaches to the multi-criteria portfolio selection are proposed, such as the use of goal programming or the use of the utility functions of distribution moments [9, 13, 15].

The importance of skewness in the rate of return distribution is introduced by Arditti and Levy in the stocks' pricing [1]. He claims that investors prefer positive skewness as a result of decreasing absolute risk aversion. The preference for positive skewness is also related to the concept of prudence introduced by Kimball [11]. A prudent investor is characterized by a special behavior in a situation of risk - precautionary saving or limiting consumption. In the context of the theory of expected utility, prudence means the convexity of marginal utility function [2, 12].

[1]University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland, renata.dudzinska-baryla@ue.katowice.pl.
[2]University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland, donata.kopanska-brodka@ue.katowice.pl.
[3]University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland, ewa.michalska @ue.katowice.pl.

The literature also explores efficient portfolios in terms of the first three moments of the rate of return distribution [1, 6]. In addition, the problem of diversification of optimal portfolios depending on the investor's preferences regarding the skewness of the rate of return is discussed. Independent research confirms that the greater the strength of preference for positive skewness, the lower the degree of portfolio diversification [7].

The goal programming (GP) has been the most widely used approach in decision-making problems with several conflicting and competing objectives. An important feature of GP is the existence of an optimal solution. Generally, the GP models vary in the form of the achievement function, which minimizes the unwanted deviations (absolute or relative) of the model's goals [13].

The polynomial goal programming (PGP) approach to the portfolio selection with skewness was proposed by Lai [15]. This article initiated the intensive development of literature, in which portfolio optimization is conceived as a multiple goal programming problem. Lai [15] and authors of other papers [4, 5, 18] assumed that the portfolio decision depends on the percentage invested in each asset and constructed a PGP model such that the portfolio choice can be rescaled and restricted on the unit variance space. This PGP methodology has become popular for empirical research looking at skewness persistence in a variety of international markets [3, 16], but under certain conditions, the PGP method can produce mathematically feasible solutions, which would be unfeasible from a financial point of view.

In our approach, the goal programming model with the weighted linear function of relative deviations from the model's goals is considered. Because of lack of general procedure describing mean-variance-skewess efficient frontier the trade-off curves in the two-dimensional space are our basic contention. The goal programming approach with the objective function depending on the strength of preference towards desired moments provides an original projection of the efficient frontier and a new interpretation of these projections. We utilize the GP to determine the trade-off frontiers of portfolios of selected stocks which constitute WIG20 index on the Warsaw Stock Exchange.

2 Mean-variance-skewness optimization problem – the goal programming approach

A portfolio is a set of securities whose shares in portfolio form a vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$, where $x_1 + x_2 + \ldots + x_N = 1$ and $x_i \geq 0$ for $i = 1, \ldots, N$. The condition for non-negativity of shares means that short sale is forbidden. Random rates of return of stocks make up a vector $\boldsymbol{R} = [R_1, R_2, \ldots, R_N]$. Rate of return of a portfolio is a random variable which distribution is generated by the random rates of return of stocks, hence $R_P = R_1 x_1 + R_2 x_2 + \ldots + R_N x_N$. Our optimization models make use of three moments of a portfolio random rate of return R_P: the expected value of the portfolio (E_P) which is the first order moment, the variance of portfolio (V_P) which is the second central moment and the skewness of portfolio (S_P) measured by the third central moment [7, 14].

Criteria for maximizing the expected value and skewness and minimizing variance are usually not met at the same time. Therefore, it is justified to consider some trade-offs between them. In our analysis of such trade-offs we propose to use the goal programming model in which deviations from the desired values of distribution moments are penalized. Each penalty expresses the strength of preference regarding the achievement of the expected rate of return, skewness, and variance, respectively.

The use of the goal programming model must be preceded by the determination of the desired levels of goals (aspiration levels). In our approach, we use the optimal values of the portfolio's expected rate of return and skewness for a given level of portfolio variance. In the first step for the assumed level of variance V_0, we optimize a well-known Markowitz portfolio selection model of the form

$$\begin{aligned} &max\, E_P \\ &V_P = V_0 \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, i = 1, \ldots, N \end{aligned} \tag{1}$$

In our procedure the meaningful values of variance has to be higher than the variance of the global minimum variance portfolio and lower than the minimum of the variances of stock with the highest expected rate of return

and stock with the highest skewness. The expected rate of return of optimal portfolio based on model (1) is noted as E_0. Pairs of values (V_0, E_0) represents points which form Markowitz's efficient frontier.

In the second step we determine the desired level of skewness. The optimal portfolio which maximizes the skewness for a given variance is a solution of the following model

$$\begin{aligned} &max\, S_P \\ &V_P = V_0 \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, i = 1, \ldots, N \end{aligned} \tag{2}$$

The skewness of optimal portfolio's rate of return based on model (2) we note as S_0. Points (V_0, S_0) can also be interpreted as efficient points in the skewness-variance space. Therefore, the values E_0 and S_0 are the most desired values of portfolio's expected rate of return and skewness for assumed level of variance V_0.

The procedure for selecting the desired levels of expected return and skewness guarantees that there is no other portfolio with higher expected return and higher skewness. Therefore, in order to analyze the trade-offs between the first three moments of rates of return, we propose the following goal programming model:

$$\begin{aligned} &min\left(\alpha\frac{de^-}{|E_0|} + \beta\frac{ds^-}{|S_0|} + \gamma\frac{dv^+}{V_0}\right) \\ &E_P + de^- = E_0 \\ &S_P + ds^- = S_0 \\ &V_P - dv^+ = V_0 \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, i = 1, \ldots, N \\ &de^-, ds^-, dv^+ \geq 0 \end{aligned} \tag{3}$$

where α, β, γ are weights fulfilling condition $\alpha + \beta + \gamma = 1$, E_0, S_0, V_0 are desired levels of goals. From the investor's point of view, deviations de^- (expected rate of return below E_0), ds^- (skewness below S_0), and dv^+ (variance above V_0) are unfavorable. The objective function of model (3) minimizes the relative unfavourable deviations from desired values. The strength of preference regarding the desired expected rate of return is reflected by weight α, the strength of preference regarding the desired skewness is reflected by weight β, and the strength of preference regarding the desired variance is reflected by weight γ. All weights have to sum up to 1.

If only one weight equals 1 (and others equal 0 at the same time) then the other moments of solution of model (3) is unpredictable. For example, if $\alpha = 1$ then the optimal portfolio is any portfolio with the expected rate of return E_0, but the skewness and variance may have any value, regardless of the desired V_0 and S_0.

In a case when α or β equals 0 (and other weights are different from 1) the optimal portfolio has the other moments on assumed levels. If $\alpha = 0$ then the optimal solution of model (3) is the same as a solution of model (2). Similarly, if $\beta = 0$ then the optimal solution of model (3) is the same as a solution of model (1). But in the case when $\gamma = 0$ then the expected rate of return and skewness of optimal solution of model (3) varies between parameters of optimal portfolios based on model (1) and (2) depending on a proportion between α and β weights.

3 Trade-offs – empirical research

The purpose of our analysis is to illustrate how selected parameters of the return rate distribution of the optimal portfolio varies for investors with different preferences for the expected value, skewness, and variance. These preferences are expressed by means of the α, β and γ parameters.

The data set contains daily logarithmic rates of return for the year 2017 for all 20 stocks from WIG20 index quoted on the Warsaw Stock Exchange. Calculations were made in the SAS software using the NLP solver and self-prepared programs.

In this paper we analyze the expected values and skewness of rates of return of optimal portfolios for various values of variance. Values of variance have to be from the interval $V_0 \in \langle V_{\text{GMVP}}, \min_i\{V_{max\,E_i}, V_{max\,S_i}\}\rangle$ where V_{GMVP} is a variance of the global minimum variance portfolio, $V_{max\,E_i}$ is a variance of a stock with the highest expected value, $V_{max\,S_i}$ is a variance of a stock with the highest skewness. In other words, the assumed variance V_0 has to be not lower than the minimal variance of all portfolios and not higher than the lower from variances of the maximum expected value stock and the maximum skewness stock. Table 1 shows the first three moments of logarithmic returns for all 20 stocks and the global minimal variance portfolio (GMVP) in the analyzed period.

No.	Stock	E	V	S	No.	Stock	E	V	S
1	ALIOR	0.153	3.363	2.887	11	LPP	0.181	3.628	-0.138
2	ASSECOPOL	-0.082	2.116	-0.011	12	MBANK	0.131	4.399	1.343
3	BZWBK	0.091	**3.225**	**2.985**	13	ORANGEPL	0.020	3.591	-0.634
4	CCC	0.135	3.627	0.971	14	PEKAO	0.012	1.923	-0.846
5	CYFRPLSAT	0.004	2.350	-0.982	15	PGE	0.057	3.500	-0.132
6	ENERGA	0.134	3.809	2.924	16	PGNIG	0.044	3.012	-0.097
7	EUROCASH	-0.158	4.101	-11.802	17	PKNORLEN	0.087	4.376	-5.085
8	JSW	0.146	6.110	2.573	18	PKOBP	**0.182**	**2.908**	2.105
9	KGHM	0.074	3.960	2.780	19	PZU	0.095	2.221	0.178
10	LOTOS	0.164	5.005	1.669	20	TAURONPE	0.027	3.086	2.655
						GMVP	0.047	**0.592**	0.093

Table 1 Expected values (E), variances (V) and skewnesses (S) of stocks and GMVP

Based on values in Table 1 the interval of the assumed variance V_0 is $\langle 0.592, 2.908\rangle$. In the optimization models (1), (2) and (3) we considered values rounded to 0.1.

In the following part, only some fragments of empirical research are presented. For selected values of parameters, trade-off lines are shown in relation to the corresponding efficient frontiers.

In Fig.1 the E0 line is the efficient frontier obtained as a solution of model (1) for different levels of V_0, i.e. the line of the highest possible expected rates of return of portfolios. Including additional preferences regarding the desired level of skewness results in that the obtained optimal portfolios may have a lower expected rate of return, and therefore do not overlap with the efficient frontier. The corresponding curves create trade-off lines in the mean-variance space. For low strength of preference regarding the desired expected rate of return ($\alpha = 0.2$), the trade-off lines overlap with the efficient frontier for the desired variance V_0 lower than 1.8. In addition, along with the increase in strength of preference regarding the achievement of the desired value of skewness, trade-off fragments move away from the efficient line and reach the boundary position for $\beta = 0.6$.

A similar relationship is observed for the strength of preference α not exceeding 0.4. For a relatively strong preference of reaching the desired expected value ($\alpha > 0.4$), trade-off lines overlap with the efficient line.

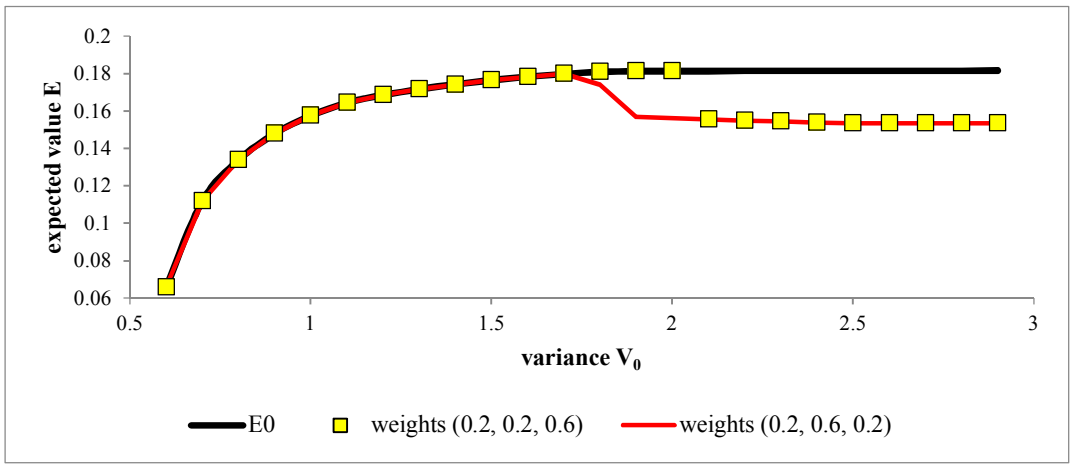


Figure 1 Optimal expected rates of return for selected preference weights (α, β, γ) and variance V_0

As it was in the case of the expected value, the S0 line is the curve of optimal skewnesses of the model (2) for different levels of V_0 (Fig. 2). For the strength of preference of reaching the desired skewness at $\beta = 0.2$,

trade-off lines overlap with the curve of optimal skewnesses for desired variance V_0 lower than 1.7. In addition, along with the increase in the strength of preference regarding the achievement of desired expected return rate, fragments of trade-off lines move away from the line of optimal skewnesses and reach the boundary position for $\alpha = 0.4$. Regardless of the strength of preference of reaching the desired third moment of the distribution, for the desired level of variance less than 1.7, the corresponding fragment of the trade-off line in the skewness-variance space overlaps the boundary curve of optimal skewnesses.

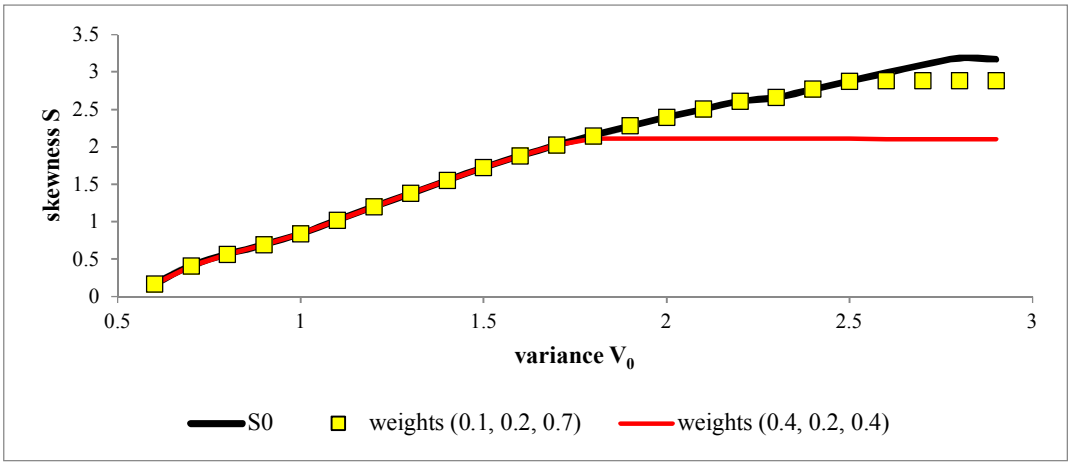


Figure 2 Optimal skewness for selected preference weights (α, β, γ) and variance V_0

The graph in Fig.3 shows the optimal variances of portfolios obtained for the desired variances (line V0). For the strength of preference regarding the desired variance at the level of $\gamma = 0.4$, deviations from the desired values increase monotonically with the increase of the level of the desired variance in the interval $\langle V_{\mathrm{GMVP}}, 1.7\rangle$. The same observation was made for other levels of strength of preference regarding variance.

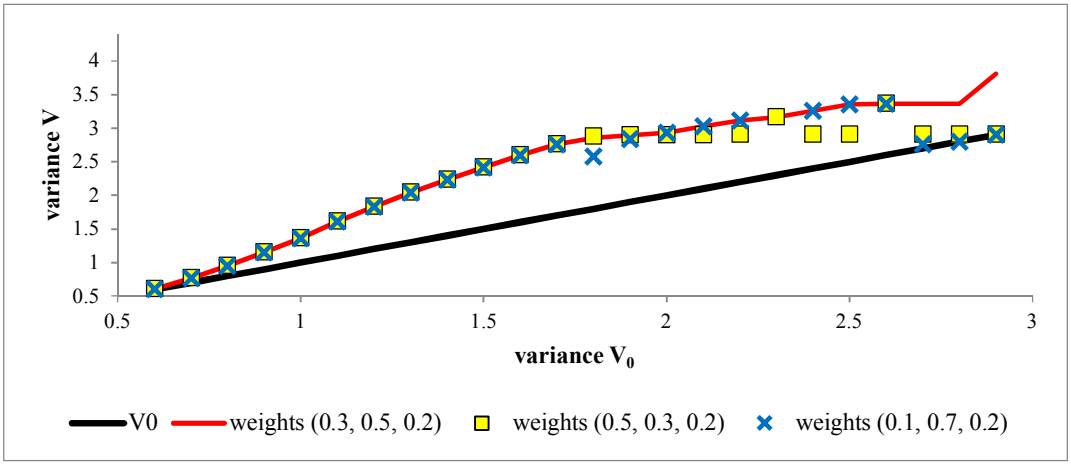


Figure 3 Optimal variance for selected preference weights (α, β, γ) and variance V_0

4 Conclusions

Our results clearly show a trade-off between expected value, variance and skewness which was traditionally assumed to be present only between expected value, and variance. The investors aware of additional criteria would have to accept a lower rate of return if they chose to include in optimization the skewness of the portfolio rate of return. This means that the observed efficient frontier based only on mean-variance optimization is not an adequate efficient frontier and could lead investors towards sub-optimal decisions when the skewness is considered.

Our study adopts the GP approach to include the third moment of the rate of return distribution in the portfolio optimization process. The investors' aim is to provide portfolios with the highest expected rate of return and the highest third central moment and the lowest variance. Because these criteria are mutually competitive, inves-

tor expresses his own preferences for each criterion relative to the others by the subjective value of the parameter measuring the preference strength. This parameter appears in the objective function in a proposed GP model.

For the observed stocks in a given period, regardless of the strength of preference towards any distribution moment, the trade-off lines of the expected rate of return and skewness are characterized in that for values of the desired variance lower than 1.7 they overlap with the relevant efficient line. In addition, deviations of variances of optimal portfolios from the desired values of goals increase monotonically for V_0 lower than 1.7. Thus, the parameters of optimal portfolios received by means of goal programming for variances lower than 1.7, have special properties.

References

[1] Arditti, F.D. & Levy, H. (1975). Portfolio Efficiency Analysis in Three Moments: The Multiperiod Case. *Journal of Finance, 30(3)*, 797–809.

[2] Briec, W., Kerstens, K. & Jokung, O. (2007). Mean Variance Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach. *Management Science, 53(1)*, 135–149.

[3] Chen, H.H. & Shia, B.C. (2007). Multinational Portfolio Construction Using Polynomial Goal Programming and Lower Partial Moments. *Journal of the Chinese Statistical Association, 45*, 130–143.

[4] Chunhachinda, P., Dandapani, K., Hamid, S. & Prakash, A.J. (1997). Portfolio Selection and Skewness: Evidence from International Stock Markets. *Journal of Banking and Finance, 21*, 143–167.

[5] Davies, R.J., Kat, H.M. & Lu, S. (2009). Fund of Hedge Funds Portfolio Selection: A Multiple-Objective Approach. *Journal of Derivatives and Hedge Funds, 15(2)*, 91–115.

[6] Dudzińska-Baryła, R., Kopańska-Bródka, D. & Michalska, E. (2015). Analiza portfeli narożnych z uwzględnieniem skośności (Analysis of Skewness for Corner Portfolios). *Finanse, Rynki Finansowe, Ubezpieczenia, 75*, 123–133 (in Polish).

[7] Dudzińska-Baryła, R., Kopańska-Bródka, D. & Michalska, E. (2017). Diversification Problem in Mean-Variance-Skewness Portfolio Models. In: P. Pražák (Ed.), *Conference Proceedings, 35rd International Conference Mathematical Methods in Economics* (pp. 137–142). Hradec Králové: University of Hradec Králové.

[8] Fama, E.F. (1965). The Behavior of Stock Market Prices. *Journal of Business, 38(1)*, 34–105.

[9] Harvey, C.R., Liechty, J.C., Liechty, M.W. & Müller, P. (2010). Portfolio Selection with Higher Moments. *Quantitative Finance, 10(5)*, 469–485.

[10] Joro , T. & Na P. (2006). Portfolio Performance Evaluation in Mean-Variance-Skewness Framework, *European Journal of Operational Research, 175(1)*, 446–461.

[11] Kimball, M.S. (1990). Precautionary Saving in the Small and in the Large. *Econometrica, 58(1)*, 53–73.

[12] Kopańska-Bródka, D. (2013). Związki miar awersji do ryzyka z funkcjami zależnymi od parametrów rozkładu. Studia Ekonomiczne (Relations of Measures of Risk Aversion with the Functions Depended on the Parameters of the Probability Distribution). *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 163*, 253–266 (in Polish).

[13] Kopańska-Bródka, D. (2014). Optymalny portfel inwestycyjny z kryterium maksymalnej skośności (Optimal Investment Portfolio for Skewness Maximization Criteria). *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 208*, 46–58 (in Polish).

[14] Kopańska-Bródka, D., Dudzińska-Baryła, R. & Michalska, E. (2019). The Investor's Preferences in the Portfolio Selection Problem Based on the Goal Programming Approach. In: W. Tarczyński, K. Nermend (Eds.) Effective Investments on Capital Markets. Series: Springer Proceedings in Business and Economics. Springer.

[15] Lai, T.Y. (1991). Portfolio Selection with Skewness: A Multiple-Objective Approach. *Review of Quantitative Finance and Accounting, 1*, 293–305.

[16] Naqvi, B., Mirza, N., Naqvi, W.A. & Rizvi, S.K.A. (2017). Portfolio Optimisation with Higher Moments of Risk at the Pakistan Stock Exchange. *Economic Research-Ekonomska Istraživanja, 30(1)*, 1594–1610.

[17] Piasecki, K. & Tomasik, E. (2013). *Rozkład stóp zwrotu z instrumentów polskiego rynku kapitałowego*. Kraków-Warszawa: Edu-Libri (in Polish).

[18] Prakash, A.J., Chang, C.H. & Pactwa, T.E. (2003). Selecting a Portfolio with Skewness: Recent Evidence from US, European, and Latin American Equity Markets. *Journal of Banking and Finance, 27*, 1375–1390.

[19] Scott, R. & Horvath, P. (1980). On the Direction of Preference for Moments of Higher Order than the Variance. *Journal of Finance, 35*, 915–919.

[20] Simonson, D. (1972). The Speculative Behavior of Mutual Funds. *Journal of Finance, 27*, 381–391.

Cost analysis of rectifying acceptance sampling plans

Nikola Kaspříková[1]

Abstract. The paper addresses the LTPD acceptance sampling plans for the inspection by variables minimizing the mean inspection cost per lot of the process average quality when the remainder of the rejected lots is inspected by attributes. The sampling plans based on the EWMA statistic may bring significant savings in the mean inspection cost per lot of the process average quality over the originally designed plans by Dodge and Romig for the inspection just by attributes. The extent of the savings is dependent, among other parameters, on the cost of inspecting an item by variables and the cost of inspecting the item by attributes. The break even value of the quotient of the cost of inspecting an item by variables and the cost of inspecting the item by attributes is calculated and analysed for the plans, comparing the situation with known standard deviation and the situation with the unknown standard deviation of the quality characteristics. This value may be used to make the decision whether to use the variables or the attributes sampling plan in a concrete situation in practice. The calculations can be performed with the free LTPDvar extension package for the R software.

Keywords: Acceptance sampling, LTPD, cost, inspection by variables.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The rectifying lot tolerance percent defective (LTPD) plans minimizing the mean inspection cost per lot of the process average quality have been originally designed by Dodge and Romig for the inspection by attributes, see [1]. Plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of rejected lots is inspected by attributes) considering just the information from the current lot were then proposed and it was shown that these plans are in many situations more economical than the corresponding Dodge-Romig attribute sampling plans, see [5]. The paper [3] has introduced the rectifying LTPD plans for the sampling inspection by variables based on the exponentially weighted moving average (EWMA) statistic. Such plans, designed in [3], actually use not only the information from the current lot, but also the information obtained from the previous lots and are a sort of plans with memory. It has been shown in [3] that in many situations in the business practice the EWMA based plans bring significant economic advantage over the corresponding attribute sampling plans and over the variable sampling plans considering just the information on the quality of the current lot. The LTPDvar software [2] which has been published on the Comprehensive R Archive Network is an extension package for R computing environment [6] developed for the computation and evaluation of optimal plans for statistical acceptance sampling. It is a free software and it may be used for the calculation of the LTPD and AOQL (average outgoing quality limit) variable sampling plans minimizing the mean inspection cost per lot of the average process quality when the remainder of the rejected lots is inspected.

The case study of the rectifying EWMA-based variable sampling LTPD plans minimizing the mean inspection cost per lot of process average quality is reported in the present paper and the discussion of the economic performance of such plans is provided, focusing on whether the standard deviation of the quality characteristic is known or unknown. The extent of the savings of variable sampling plan over the attribute sampling plan is dependent, among other parameters, on the cost of inspecting an item by variables and the cost of inspecting the item by attributes. The break-even value of the quotient of the cost of inspecting an item by variables and the cost of inspecting the item by attributes is calculated and analysed for the plans, comparing the situation with known standard deviation and the situation with the unknown standard deviation of the quality characteristics. Following the analysis obtained for the AOQL plans in [4], a simple economic model is used in this paper to evaluate the economic efficiency of the EWMA-based LTPD plans in comparison with the corresponding attribute sampling plans.

[1] University of Economics in Prague, Department of Mathematics,
Nám. W. Churchilla 4, Praha, Czech Republic, nb33@tulipany.cz

The structure of the paper is as follows: the LTPD plans for the inspection by attributes minimizing the mean inspection cost per lot of the process average quality when the remainder of the rejected lots is inspected are recalled first, then the rectifying LTPD plans for the inspection by variables and attributes are recalled and the evaluation of the comparative economic efficiency is shown, using a simple economic model.

2 LTPD attribute sampling plans

First we recall the attribute sampling plans, as introduced by Dodge and Romig, see e. g. [1]. Dodge and Romig consider the case, where each inspected item is to be classified as either good or non-conforming (acceptance sampling by attributes) and they search for the sampling plans which minimize the mean number of the items inspected per lot of the process average quality

$$I_s = N - (N-n)\cdot L(\bar{p};n;c) \tag{1}$$

under the condition

$$L(p_t;n;c) \leq \beta, \tag{2}$$

where $L(p,n,c)$ is the operating characteristic (the probability of accepting a submitted lot with the proportion defective p when using the plan (n,c) for acceptance sampling),
N is the number of the items in the lot (an input parameter),
$\bar{p}$ is the process average proportion defective (an input parameter),
p_t is the lot tolerance proportion defective (an input parameter, $P_t = 100p_t$ is the lot tolerance per cent defective, denoted LTPD),
n is the number of items in the sample $(n < N)$,
c is the acceptance number (the lot is rejected when the number of defective items in the sample is greater than c).

Condition (2) provides a guarantee for the consumer that the lots of the unsatisfactory quality level, with the proportion defective p_t are going to be accepted only with the specified probability β at most (the so-called consumer's risk). The standard value 0.1 is used for the consumer's risk in [1].

3 Sampling plans for the inspection by variables

We recall the LTPD plans for the inspection by variables and attributes designed in [3]. The plans have been designed under the following assumptions: The measurements of a single quality characteristic X are independent and identically distributed normal random variables with parameters μ and σ^2. We consider the unknown σ case first. For the quality characteristic X, either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit L (the item is defective if its measurement is smaller than L), is given.

For the design of the rectifying LTPD plans minimizing the mean inspection cost per lot of the process average quality we shall use the following procedure: draw a random sample of n items from the lot and compute the sample mean $\bar{x}$ and the sample standard deviation s. Then calculate the EWMA statistic Z at time t as

$$Z_t = \lambda\bar{x} + (1-\lambda)Z_{t-1}, \tag{3}$$

where λ is the smoothing constant with value between 0 and 1.

Then accept the lot if

$$\frac{U - Z_t}{s} \geq k, \quad \text{or} \quad \frac{Z_t - L}{s} \geq k. \tag{4}$$

In case that the σ value is known, we its value is used in place of s in (4). The plan parameters (n,k) are to be determined so that the plan has optimal economic characteristics and satisfies the requirement (2), using the operating characteristic reported in [3].

Regarding the economic optimality, Klufa in [5] uses the economic model, described in more detail in [4] and searches for the acceptance plan (n,k), minimizing the mean inspection cost per lot of the process average quality C_{ms} under the condition (2). We are going to use this model for the design of the plans too. The inspection cost per lot, assuming that the remainder of the rejected lots is inspected by attributes (the inspection by variables and attributes), is nc_m^*, with the probability $L(p;n,k)$, and $[nc_m^* + (N-n)c_s^*]$ with the probability $[1-L(p;n,k)]$, where c_s^* is the cost of the inspection of one item by attributes, and c_m^* is the cost of the inspection of one item by variables. The mean inspection cost per lot of the process average quality is then

$$C_{ms} = n \cdot c_m^* + (N-n) \cdot c_s^* \cdot [1 - L(\bar{p}; n, k)]. \tag{5}$$

Let us denote

$$c_m = \frac{c_m^*}{c_s^*}. \tag{6}$$

Instead of C_{ms} we shall look for the acceptance plan (n,k) minimizing

$$I_{ms} = n \cdot c_m + (N-n) \cdot [1 - L(\bar{p}; n, k)] \tag{7}$$

(both functions C_{ms} and I_{ms} have a minimum for the same acceptance plan ($C_{ms} = I_{ms} \cdot c_s^*$)) under the condition (2).

For the discussion of the c_m parameter, see [4]. The value of this parameter must be estimated based on the real cost calculation in practice. Usually it is $c_m > 1$.

For the economic comparison of the variable sampling plan and the corresponding attribute sampling plan we will use the following expression

$$e = \frac{I_{ms}}{I_s}. \tag{8}$$

For the comparison of the variable sampling plan (n_1,k) and the attribute sampling plan (n_2,c) we have

$$e = \frac{n_1 \cdot c_m + (N-n_1) \cdot (1 - L(\bar{p}, n_1, k))}{N - (N-n_2) \cdot L(\bar{p}, n_2, c)}. \tag{9}$$

The expression $(1-e) \cdot 100\%$ then shows the savings in percent in the mean inspection cost per lot of the process average quality when the variable sampling plan (n_1,k) is used in place of the corresponding attribute sampling plan (n_2,c).

The extend if the savings is dependent on the value of the c_m parameter. We define c_m^{BE} as such value of the c_m parameter for which it is

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

4 Calculation and evaluation of the plans

The LTPD acceptance sampling plan for the sampling inspection by variables when the remainder of rejected lots is inspected by attributes, as implemented in [2] will be calculated in the example below. The solution for the known σ will be searched for first, then the unknown σ case will be considered, using the operating characteristic reported in [3]. The resulting sampling plan will then be evaluated with regard to the economic characteristics and compared with the corresponding LTPD attributes sampling plan as discussed in [1], and the c_m^{BE} value will be calculated.

Example 1. A lot with $N = 1000$ items is considered in the acceptance procedure. The lot tolerance proportion defective is given to be $p_t = 0.01$, and the process average quality is $\bar{p} = 0.001$. It is known that the cost of the inspection of one item by variables is 5 times higher than the cost of the inspection of one item by attributes, so the parameter c_m equals 5. Find the EWMa-based LTPD acceptance sampling plan for sampling inspection by variables when remainder of rejected lots is inspected by attributes with $\lambda = 0.9$.

The plan can be calculated using the functions available in the LTPDvar package for the R software [6], see the documentation of the package for a more detailed description. The implementation is based on the following principles: the k corresponding to particular sample size n is obtained using the condition (2) and the cost-optimizing n is searched for using numerical methods, where the approximate solution, which is calculated first, is used for setting the bounds of the intervals to be searched in.

First we cosider the situation with known σ. The resulting optimal sampling plan is

$$n = 16,$$

$$k = 2.61615.$$

Remark 1. The optimal plan may be calculated with the LTPDvar software [2] using the function planLTPD.

One can see that the sample size for the variable sampling plan under the assumption that σ is known is rather small. For the values of the input parameters in our example, the attributes sampling plan ($n = 205$, $c = 0$) is given by [1]. Let us compare the two plans ($n = 16$, $k = 2.61615$) and ($n = 205$, $c = 0$) from economic point of view.

Let us denote the variable sampling plan (n_1, k) and the attribute sampling plan (n_2, c).

The economic efficiency of the EWMA-based variable sampling plan will be evaluated using the comparison of the mean inspection cost per lot of the process average quality of this plan and attribute sampling plan as reported by [1], more specifically, we will use the characteristic (9).

Comparing the plan $(n_1 = 16, k = 2.61615)$ and (n_2, c) = (205, 0) we get from (9)

$$e = 0.266.$$

The expression $(1 - e) \cdot 100\%$ then represents the percentage of savings in the mean inspection cost when using the variable plan $(n_1 = 16, k = 2.61615)$ in comparison with the situation, when the attribute sampling plan (n_2, c) = (205, 0) is used.

One may expect that using the variable sampling plan based on EWMA statistic in place of the corresponding attribute sampling plan allows obtaining 70 % savings of the cost.

Let us further consider another situation, with higher value of the c_m input parameter. Let us have c_m=10. For such high value of the c_m parameter, we get the optimum variable sampling plan

$$n = 13,$$

$$k = 2.647854.$$

For the plan $(n_1 = 13, k = 2.647854)$ we get from (9)

$$e = 0.458,$$

that is the result suggesting a smaller extent of the savings in comparison with the situation discussed above for c_m=5.

In case that the value of the cost parameter c_m were even higher, one gets the higher value of e. For example the resulting variable sampling plan for c_m=20 leads to

$$e = 0.765.$$

In case that c_m =25, we get

$$e = 0.895.$$

In case that c_m =29, one gets the plan $(n_1 = 8, k = 2.736189)$ and it is

$$e = 0.993.$$

For $c_m > 29.4$, the variables sampling plan is less economical than the corresponding attribute sampling plan considering the mean inspection cost, since it is $e > 1$.

The results of the calculations discussed above are summarized in Table 1.

c_m	$e \cdot 100$
5	26.6
10	45.8
20	76.5
25	89.5
29	99.3

Table 1 The impact of c_m

Keeping the other input parameters unchanged, the extent of the savings in the inspection cost $(1 - e) \cdot 100\%$ is dependent on the value of the c_m parameter, that is the quotient of the cost of inspecting the item by variables

and the cost of inspecting the item by attributes, and it is higher for the smaller values of c_m, see Table 1. In our example, it is

$$c_m^{BE} = 29.4.$$

As it has been already noted in the text above, the variable sampling plans in the known σ case are very economic with respect to the sample size. The mean inspection cost per lot of the process average quality is also quite low in comparison with the mean inspection cost per lot of the process average quality for the attribute sampling plan, considering the values of the c_m parameter near the value in our example. The c_m^{BE} value calculated for our example was rather high. In real situations in practice, such a high value is probably not going to be reached by the c_m parameter value.

For further analysis, let us now drop the assumption of the known σ parameter value of the quality charactetistic.

For the input parameters and c_m=5 we get the following variable sampling plan calculated under the unknown σ value assumption

$$n = 51,$$
$$k = 2.724137.$$

It may be noticed that the sample size for the variable sampling plan calculated for the unknown σ value of the quality characteristic is considerably smaller than the sample size for the corresponding attribute sampling plan, but it is larger than the sample size of the variable sapling plan calculated using the known σ value assumption.

We evaluate the economic performance of the variable sampling plan using the comparison of the mean inspection cost per lot of the process average quality with the cost of the corresponding attribute sampling plan (205, 0).

For c_m=5 and the plan (51, 2.724137) we get from (9)

$$e = 0.958.$$

That is the savings in the mean inspection cost per lot of the process average quality for the variable sampling plan is approximately 4 % in comparison with the attribute sampling plan.

For c_m=6 we get the variable sampling plan (47, 2.744052) and from (9) it is $e = 1.091$, that is the attribute sampling would be more economical in this case. On the other hand for smaller values of the c_m parameter than what was the value in the original assignment, we obtain variable sampling plans, for which it is $e < 1$ and the variable sampling is more economical with respect to the mean inspection cost than the attribute sampling. The results of the calculation of e in response to the value of the c_m parameter may be summarized in Table 2.

c_m	$e \cdot 100$
2	47,7
3	65,4
4	81,3
5	95,8
6	109,1

Table 2 The impact of c_m

Using the variable sampling plans is obviously considerably more economical than using the attribute sampling plans, even though the situation when the σ parameter value is unknown brings smaller advantage in comparison with the known σ case. For example if it is c_m=2, the savings in the mean cost of inspection when using the variable sampling plan is over 50 %. In our example for the unknown σ it is

$$c_m^{BE} = 5.3.$$

That means if the real value of the c_m parameter is below 5.3, it is more economical to use the variable sampling plan.

For the case considered in our example, the value of e in response to the values of the input cost parameter c_m is shown in Figure 1, which shown also the value of c_m^{BE}.

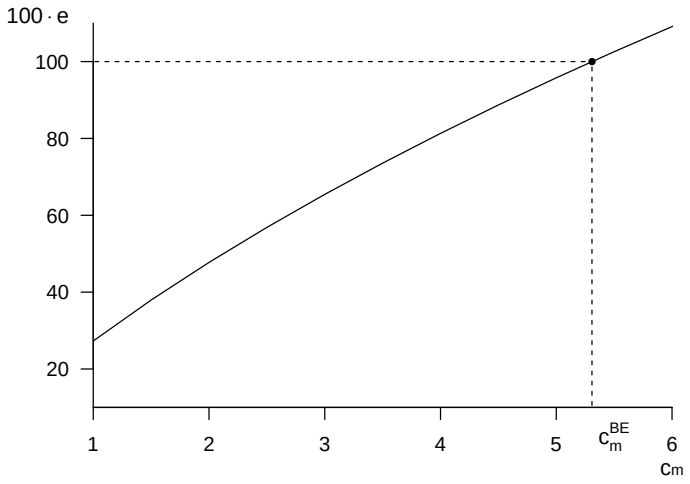


Figure 1 Economic efficiency, comparison with plan (205, 0)

5 Conclusions

Considering the mean inspection cost per lot of the process average quality, it has been shown that the EWMA-based rectifying LTPD variable sampling plans are in many situations more economical than the corresponding attribute sampling plans, with greater benefit in case that the σ parameter of the quality characteristic is known. This is also reflected in the calculated value c_m^{BE}, the break-even value of the c_m parameter, which is higher in the known σ case.

References

[1] Dodge, H., F. & Romig, H. G. (1998). *Sampling Inspection Tables: Single and Double Sampling*. New York: John Wiley.

[2] Kasprikova, N. (2016). *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.2. http://CRAN.R-project.org/package=LTPDvar.

[3] Kasprikova, N. (2017). Rectifying LTPD Plans Minimizing Mean Inspection Cost Per Lot of Process Average Quality Using EWMA Statistics. *Journal of Testing and Evaluation* **4**, DOI: 10.1520/JTE20150259.

[4] Kasprikova, N. & Klufa, J. (2015). AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes, *Quality technology and quantitative management* **6**, 133–142.

[5] Klufa, J. (2010). Exact calculation of the Dodge-Romig LTPD single sampling plans for inspection by variables. *Statistical Papers*, 51 (2), pp. 297-305.

[6] R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Optimality Conditions in Behavior Optimization Model of Consumers in the Network Industries Markets

Michal Fendek[1], Eleonora Fendeková[2]

Abstract. Equilibrium on the network industries market, as well as on any market, is being created based on the level of demand and supply on relevant market. Currently a significant attention in scholarly discussions on various levels is being paid to the subject of network industries. It is understandable as network industries in fact ensure the production and distribution of energy sources that play a key role in an effective operation of the developed economies.
We will point out certain features of network industries market where the consumer usually is not able to substitute a product of network industry with other product of appropriate characteristics in a short time period, thus considering the product being exclusive.
This exclusivity can be formally represented in the utility function and other related analytical tasks. In this paper, we will discuss the analysis of microeconomic optimization models of consumers and producers behavior on the network industries market, i.e. the analysis of demand and supply phenomena on this specific market.
For the optimization problems, we will formulate the Kuhn-Tucker optimality conditions and we will study their interpretation options.

Keywords: Network industries market, utility maximization model, optimality conditions, competitive environment.

JEL Classification: D42, D43, L13
AMS Classification: 49M29, 49M37, 91B16

1 Introduction

For current optimization pricing strategies in the network industries, it is typical that traditional pricing mechanisms corresponding to the now-classifying scheme of a perfectly competitive market environment are replaced by adequate mechanisms for monopoly and oligopolistic market structures, which are enriched by analytical regulation schemes in case of network operators.

Fendekova & Fendek [7] show, that for the network industries market equilibrium models a certain segregation of the market is characteristic, resulting in the network industries production usually not being substitutable. Therefore a satisfaction gained from their consumption can be uniquely quantified. In principle, it is such a representation of utility function where a network industry product is considered to be a good with its own and exactly formulated utility function and the other goods are considered as a consumption of one calculated aggregated good with standardized unit price.

Surplus of a consumer or a producer as the standard categories of microeconomic analysis can be very effectively used to describe the behavior of the consumers and producers on the market of goods and services, as they allow us to competently explain an intuitive tendencies of, mostly, consumers in creating their consumer strategy regarding the changing marketplace parameters, in this case mostly prices.

In this article we point out a specific interpretation of "consumer surplus" in case of network industries production demand analysis. Its specifics result from the fact that a consumer is usually not able to substitute network indus-

[1] University of Economics Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: michal.fendek@euba.sk.
[2] University of Economics Bratislava, Department of Business Economics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: eleonora.fendekova@euba.sk.

tries production, e.g. gas, for other goods with adequate performance, thus seeing it as exclusive and this exclusivity can be then formally interpreted in the construction of the utility function and other resulting analytical problems.

Suppose there is m consumers S_i for $i = 1, 2... m$ on the relevant network industry market. The network industry good or service of a homogenous character, let it be electricity distribution, is provided by n subjects, suppliers D_j for $j = 1, 2\ldots n$. We examine market of the homogenous good where a consumption of the good in the volume of x_i is a consumption of the homogenous good of the i-th consumer S_i and a consumption of all the other goods in a market basket of this consumer is represented by calculated aggregate variable x_{0i}. If a utility of consuming the network industry product for a consumer S_i is given by $u_i(x_i)$, which represents utility in monetary units, and the price of the calculated good is standardized to "1" then total utility of consumer S_i in monetary units is represented by the function $v_i(x_i, x_{0i})$ as follows

$$v_i(x_i, x_{0i}) = u_i(x_i) + x_{0i}$$

where

$$v_i(x_i, x_{0i}): R^2 \to R$$
$$u_i(x_i): R \to R$$

This perception of the utility function allows us to interpret it as a total utility in "monetary units" of a consumer buying x_i units of the network industry product and concurrently buying x_{0i} units of aggregate other goods of consumer basket which are priced by standardized price of one monetary unit.

Notice, that in result of utility function $v_i(x_i,x_{0i})$ being linear to variable x_{0i} – demand for the aggregated goods, marginal degree of substitution does not depend either on consumption of this aggregated good nor on its utility as

$$\frac{\partial v_i(x_i, x_{0i})}{\partial x_i} dx_i + \frac{\partial v_i(x_i, x_{0i})}{\partial x_{0i}} dx_{0i} = 0$$

$$-\frac{dx_{0i}}{dx_i} = \frac{\dfrac{\partial(u_i(x_i) + x_{0i})}{\partial x_i}}{\dfrac{\partial(u_i(x_i) + x_{0i})}{\partial x_{0i}}} = \frac{\dfrac{\partial(u_i(x_i) + x_{0i})}{\partial x_i}}{1} = \frac{\partial(u_i(x_i) + x_{0i})}{\partial x_i} = \frac{\partial u_i(x_i)}{\partial x_i}$$

$$-\frac{dx_{0i}}{dx_i} = u_i'(x_i) = mu_i(x_i)$$

where

$$u_i'(x_i) = mu_i(x_i): R \to R$$

is a function of marginal utility of consumption of the network industry product of the i-th consumer. It can then be said that a consumer compensates e.g. decrease of 1 percent in the network industry product consumption by increasing his consumption of the „aggregated" good by number of percentage equal to marginal utility of the network industry product.

Suppose the utility function of the network industry product $u_i(x_i)$ is for each consumer S_i smooth, i.e. continuous and differentiable and its value for zero consumption is also zero $u_i(0) = 0$. Also suppose the utility function is concave. Then suppose the marginal utility function $mu_i(x_i)$ is decreasing, therefore for second derivative of utility function $u_i''(x_i)$ follows that

$$u_i''(x_i) = \frac{dmu_i(x_i)}{dx_i} = \frac{d^2 u_i(x_i)}{dx_i^2} < 0$$

while the domain of marginal utility $mu_i(x_i)$ and its range follow

$$D(mu_i(x_i)) = \langle 0, \infty), \quad H(mu_i(x_i)) = (-\infty, \infty).$$

In other words, marginal utility of said network industry product is positive in its growing region but its values are gradually decreasing so that in the point of utility function maxima reaches marginal utility function zero value and at further consumption growth the decrease of marginal utility below zero is possible. Utility $u_i(x_i)$ for purchase

of x_i units of the network industry product then represents the willingness of a consumer "to pay" for x_i units of said good with adequate units x_{0i} of aggregated good. Simply said, a consumer is ready to give up as much as $u_i(x_i)$ units of aggregated good with standardized unitary price in order to gain x_i units of the network industry product while maintaining the same level of utility.

2 Optimality Conditions of the Consumers for the Network Industries Market

Behavior of the i-th consumer S_i for every i=1, 2, ..., m will be examined through optimization problem of total utility function maximization of the i-th consumer S_i with limited consumption expenditures with the limit w_i and a price of the network industry product p. This problem is for strictly positive variables x_i a x_{0i} represented:

$$v_i(x_i, x_{0i}) = u_i(x_i) + x_{0i} \quad \rightarrow \max$$

subject to

$$px_i + x_{0i} = w_i$$
$$x_i, x_{0i} \geq 0$$

In [11] we can see, that, the above stated mathematical programming optimization problem is a maximization problem of bounded extrema. Let us modify this problem to standardized form i.e. minimization problem:

$$-v_i(x_i, x_{0i}) = -u_i(x_i) - x_{0i} \quad \rightarrow \min \tag{1}$$

subject to

$$px_i + x_{0i} = w_i \tag{2}$$
$$x_i, x_{0i} \geq 0 \tag{3}$$

For this problem we will formulate generalized Lagrange function. Note that generalized Lagrange function does not explicitly include conditions of the variables being strictly positive, but they are included implicitly in Kuhn-Tucker optimality conditions. Bazaraa & Shetty [3] show that the generalized Lagrange function of the mathematical programming problem (1)... (3) is:

$$\begin{aligned} L_i(x_i, x_{0i}, \lambda_i) &= -v_i(x_i, x_{0i}) + \lambda_i(px_i + x_{0i} - w_i) = \\ &-u_i(x_i) - x_{0i} + \lambda_i(px_i + x_{0i} - w_i) \end{aligned} \tag{4}$$

We formulate the Kuhn-Tucker optimality conditions for the Lagrange function (4) for the i-th consumer S_i as follows

$$\frac{\partial L_i(x_i, x_{0i}, \lambda_i)}{\partial x_i} \geq 0 \qquad \frac{\partial L_i(x_i, x_{0i}, \lambda_i)}{\partial x_{0i}} \geq 0 \qquad \frac{\partial L_i(x_i, x_{0i}, \lambda_i)}{\partial \lambda_i} = 0$$
$$x_i \frac{\partial L_i(x_i, x_{0i}, \lambda_i)}{\partial x_i} = 0 \qquad x_{0i} \frac{\partial L_i(x_i, x_{0i}, \lambda_i)}{\partial x_{0i}} = 0 \tag{5}$$
$$x_i \geq 0 \qquad x_{0i} \geq 0$$

The Kuhn-Tucker optimality conditions (5) after using the analytical form of the Lagrange (4) and other updates we get:

$$\frac{-\partial v_i(x_i, x_{0i})}{\partial x_i} + \lambda_i p \geq 0 \qquad \frac{-\partial v_i(x_i, x_{0i})}{\partial x_{0i}} + \frac{\partial \lambda_i x_{0i}}{\partial x_{0i}} \geq 0 \qquad \frac{\partial \lambda_i (px_i + x_{0i} - w_i)_i}{\partial \lambda_i} = 0$$
$$x_i \left(\frac{-\partial v_i(x_i, x_{0i})}{\partial x_i} + \lambda_i p \right) = 0 \qquad x_{0i} \left(\frac{-\partial v_i(x_i, x_{0i})}{\partial x_{0i}} + \frac{\partial \lambda_i x_{0i}}{\partial x_{0i}} \right) = 0$$
$$x_i \geq 0 \qquad x_{0i} \geq 0$$

or

$$-\frac{\partial u_i(x_i)}{\partial x_i}+\lambda_i p\geq 0 \qquad -1+\lambda_i\geq 0 \qquad \left(px_i+x_{0i}-w_i\right)=0$$

$$x_i\left(-\frac{\partial u_i(x_i)}{\partial x_i}+\lambda_i p\right)=0 \qquad x_{0i}\left(-1+\lambda_i\right)=0$$

$$x_i\geq 0 \qquad x_{0i}\geq 0$$

And after other updates we finally get this form of the Kuhn-Tucker optimality conditions:

$$\begin{array}{llllll} \lambda_i p-mu_i(x_i)\geq 0 & \text{(a)} & \lambda_i-1\geq 0 & \text{(d)} & px_i+x_{0i}=w_i & \text{(g)} \\ x_i\left(\lambda_i p-mu_i(x_i)\right)=0 & \text{(b)} & x_{0i}\left(\lambda_i-1\right)=0 & \text{(e)} & & \text{(6)} \\ x_i\geq 0 & \text{(c)} & x_{0i}\geq 0 & \text{(f)} & & \end{array}$$

In other words if a consumer is willing to identify the optimal consumer strategy (x_i^*, x_{0i}^*), that means that the consumption of x_i^* units of the network industry product with a price p and the consumption of x_{0i}^* units of remaining aggregated goods with price 1 maximize his total utility $v\,(x_i^*, x_{0i}^*)$, multiplier λ_i^* must exist, for which the Kuhn-Tucker optimality conditions (6) hold, thus variables vector (x_i^*, x_{0i}^* , λ_i^*) is a solution to (a), (b), ..., (g).

We now point out certain interesting economically interpretable consequences of the Kuhn-Tucker optimality conditions in the context of consumer behavior analysis on the network industries market:

1. Fendek & Fendekova [9] show, that condition (g) implies that optimum market basket of the *i*-th consumer (x_i^*, x_{0i}^*) at a price p of the network industry product and the price of aggregated good being 1can be purchased for customers budget w_i.
2. Condition (e) implies that for positive optimum consumption of the aggregated good x_{0i}^* the optimum value of the Lagrange multiplier is 1, i.e. $\lambda_i^*=1$.
3. Condition (b) then implies that at positive consumption of the aggregated good x_{0i}^* and at positive consumption of the network industry product x_i^* which maximizes utility necessarily holds that in the point of maximum utility of the good its marginal utility equals its price, since Lagrange multiplier equals one $\lambda_i^*=1$ and holds
4. $$mu_i\left(x_i^*\right)=\left[\frac{du(x_i)}{dx_i}\right]_{x_i=x_i^*}=p \qquad (7)$$
5. Consequence (3) also confirms another important theoretical postulate, namely that a consumer increases his consumption, of the network industry product in this case, until the marginal utility reaches the level of good's market price.
   This statement actually also indirectly follows from the Kuhn-Tucker optimality conditions (a), (b), and we see that in case the condition (a) is realized for optimal structure of the vector (x_i^*, x_{0i}^* , λ_i^*) as strict inequality
   $$-mu_i(x_i)+p>0$$
   and marginal utility of the network industry product is therefore lover than its market price, then from (b) implies, that the consumer does not purchase any network industry product and
   $$x_i^*=0.$$
   The conclusion is logical and fully economically justifiable as in this case the increase in utility caused by the good purchase would not even cover its price.
6. Examine more closely the specific situation that would occur in case the optimal value of the Lagrange multiplier is higher than one, i.e. $\lambda_i^*>1$. But if conditions (d) and (e) hold then the consumer is not buying any aggregated good and $x_{0i}^*=0$. The condition (g) also implies that the consumer invests all his funds w to purchase the network industry product
   $$x_i^*=\frac{w}{p_i}$$

At the same time the condition (b) implies an interesting statement that the marginal utility of the last bought unit of the network industry product is still higher than its market price while the condition (a) holds as an inequality and

$$\begin{aligned} &\lambda_i = 1+\varepsilon,\ \ \varepsilon > 0 \\ &\left(-mu_i\left(x_i\right)+\left(1+\varepsilon\right)p\right)=0 \\ &mu_i\left(x_i\right)=\left(1+\varepsilon\right)p \\ &p < mu_i\left(x_i\right) \end{aligned}$$

We can see that in this case the consumer's decision is clearly effective since the purchase of "the last" unit of the network industry product brings higher increase in utility than its market price is, so in this case the consumer invests all his funds defined for this market basket to purchase the network industry product.

From the analysis of the Kuhn-Tucker optimality conditions for solving the optimization problem (1), ..., (3) implies an interesting interpretation of the relation between price p of the network industry product and optimal level of demand for the network industry product x_i^*. In condition a customer decides for positive volume of the aggregated good purchase $x_{0i}^* > 0$, from the conditions (b), (e) implies the identical conclusion as from the relation (7), which is the function of marginal utility $mu_i(x_i)$ determines corresponding supply volume for each level of market price and therefore de facto defines the demand function $d_i(p)$ of the i-th consumer S_i as follows

$$\begin{aligned} &p = mu_i\left(x_i^*\right) \\ &x_i^* = d_i\left(p\right) = mu_i^{-1}\left(p\right) \end{aligned} \tag{8}$$

As the marginal utility function is decreasing with increasing demand, the lover the price is, the higher the demand for the good.

3 Conclusion

In this article we examined the optimality conditions for the model of the consumers and producers behavior on the network industry market. We showed how the findings resulting from the Kuhn-Tucker optimality conditions analysis formulated for the relevant problems of mathematical programming can be effectively used to interpret the factual relations, principles and strategic decisions in consumers' behavior on the network industry market.

Breaking of the equilibrium conditions can of course not be eliminated, the everyday economic life even expects some development and instability on every market and thus also the relative momentariness of the conditions validity, which in fact is not an unsolvable problem. It is however important to identify the situation competently and evaluate the possible reactions to the changed parameters of the system.

If the abovementioned situations wouldn't occur, certain consumers could increase their utility by an exchange. In that case a consumer with a higher marginal utility and thus a higher marginal readiness to pay for the aggregated good could gain a corresponding amount of the network industry product for an adequate volume of the aggregated good from a consumer with a lower marginal utility and therefore both consumers would better their market positions.

Also the consequence (4) about the equality of the consumers marginal utility and producers marginal costs at the real market equilibrium is mostly momentary. In case of breaking the equilibrium, for example higher readiness of the consumers to pay over lower marginal costs of the producers, the total welfare can be increased.

Therefore we showed that the use of the model approach and the optimization theory at the network industries market supply and demand equilibrium conditions analysis allows us to effectively study the equality conditions as well as the consequences of the market parameters change which results in reassessment of equilibrium attributes.

Acknowledgements

Supported by the grant No. 1/0368/18 "*Pricing Strategies in the environment of an effective regulated mechanism in the network industries markets of the Slovak economy"* of the Grant Agency VEGA, Slovakia. (Scientific Grant Agency of the Ministry of Education, science, research and sport of the Slovak Republic and the Slovak Academy of Sciences).

References

[1] Blum, U., Müller, S. & Weiske von Gabler, A. (2006). Angewandte Industrieökonomik. Theorien - Modelle – Anwendungen. Berlin: Verlag Gabler.

[2] Bodea, T. & Ferguson, M. (2015). Segmentation, Revenue Management, and Pricing Analytics. New York and London: Routledge.

[3] Bazaraa, M. & Shetty, C.M. (2006). Nonlinear Programming: Theory and Algorithms. Wiley-Interscience. New York.

[4] Carlton, D.W. & Perloff, J. M. (2005). Modern Industrial Organization. Boston: Addison Wesley, 2005. ISBN 0-321-22341-1.

[5] Fendek, M. & Fendeková, E. (2008). Mikroekonomická analýza. IURA Edition. Bratislava 2008. 557 strán. ISBN 80-88715-54-7.

[6] Fendek, M. & Fendeková, E. (2010). Modely cenovej regulácie sieťových odvetví. Ekonomický časopis: časopis pre ekonomickú teóriu a hospodársku politiku, spoločensko-ekonomické prognózovanie. Bratislava: Ekonomický ústav SAV: Prognostický ústav SAV, ISSN 0013-3035, 2010, roč. 58, č. 10, pp. 1039 – 1055..

[7] Fendeková, E. & Fendek, M. (2018). Mikroekonómia: Oligopoly a regulované monopoly. Recenzenti: Miroslav Grznár, Marián Goga. 1. vydanie. Bratislava: Wolters Kluwer, 2018. 230 s. Ekonómia. ISBN 978-80-8168-765-5.

[8] Fendeková, E. & Fendek, M. (2015). Mikroekonomická analýza efektov cenovej diferenciácie produkcie monopolu. In Modely rovnováhy v podmienkach produktovej a cenovej diferenciácie na regulovaných trhoch sieťových odvetví / Eleonora Fendeková a kolektív. Bratislava: Vydavateľstvo EKONÓM, 2015. - ISBN 978-80-225-4035-3.

[9] Fendek, M. & Fendeková, E. (2009). Models of regulation of network industries in Slovakia. In International journal of economics and business research. - UK : Inderscience Enterprises Ltd, ISSN 1756-9850, 2009, vol. 1, no. 4, pp. 479-495.

[10] Fendek, M. (2008). Natural monopoly cost-oriented price regulation. (2008). Quantitative methods in economics: multiple criteria decision making XIV. – Bratislava: IURA EDITION, 2008. ISBN 978-80-8078-217-7, pp. 45-53.

[11] Jarre, F. & Stoer, J. (2004). Optimierung. Berlin: Springer Verlag. ISBN3-540-43575-1.

[12] Martin, S. (2012). Industrial Organization: A European Perspective. Oxford: Oxford University Press, USA.

[13] Shy, Oz. (2001). The Economics of Network Industries. Cambridge: Cambridge University Press.

[14] Shy, Oz. (2008). How to Price. A Guide to Pricing Technics and Yeld Management. New York: Cambridge Univesity Press.

An alternative approach towards dealing with uncertainty in project time analysis

Robert Hlavatý[1], Helena Brožová[2]

Abstract. Critical path method is a fundamental tool of time analysis in project planning. Due to its deterministic nature, it does not reflect any aspects of uncertainty that might occur in the actual real-world applications. The ways of embedding uncertainty into mathematical model of project time analysis have been widely used, in its basic form represented by probabilistic approach of PERT or GERT. We introduce an alternative viewpoint on uncertainty appearing in activity duration evaluation. First, the stated problem is formulated as a longest path problem in directed acyclic graph in the form of mixed integer linear program. Further on, we take an advantage of two different robust formulations that allow to identify critical scenarios in case any deviations from deterministic values of activity durations should appear. It turns out that different concepts of robustness must be used depending whether the duration of an activity was prolonged or contracted. The resulting scenarios show the worst-case situations with the deviations considered while it is possible to identify those activities that bear somewhat higher potential of criticalness towards the total project duration.

Keywords: Critical path method, criticalness, robust optimization, worst-case scenario

JEL Classification: C61
AMS Classification: 90B10, 90C08, 90C11

1 Introduction

The critical path method introduced by Kelly and Walker [12] has been a foundation for project planning time analysis for 60 years now. Yet, it is still a focus of many researchers who try to create extended formulations of the problem to capture the reality in a better way and support decision makers in their strategic thinking. The uncertainty in project planning is often understood as a negative phenomenon, however, the authors in the field point out that the uncertainty can be considered an opportunity as well [10,13]. Böhle, Heidling and Schoper [4] mention that the project managers often try to avoid the uncertainties in the first place rather than using them to their advantage. That is why there is a need for dealing with the uncertainty in a proper way. There are three approaches of handling uncertainty – stochastic, fuzzy and robust.

There is a review elaborated by Herroelen and Leus [7] that summarizes either of these techniques in one place, even if a bit outdated now. The authors [7] see the potential in stochastic and robust approach while the fuzzy approach is rather critized. Since our paper is focused on the robust approach, only the works in that field are discussed in the review here. Chakrabortty, Sarker and Essam [6] review the robust approach in relation to resource constraints depending on the type of uncertainty considered in durations. Another paper of Bruni et al. [5] discuss the same approach to robust modelling as we do and propose a similar model that is the same in principles, however, not considering the possible contraction of activity duration. The robust approach has also found its use in practically oriented papers, like the one of Ke, Wang and Huang [11] dedicated to scheduling in maritime logistics. Finally, there is a paper that tries to span the stochastic, fuzzy and robust approach by Sadjadi, Pourmoayed and Aryanezhad [15] or a work of Na, Wuliang and Hua [14] who capture the aspect of robustness by Monte Carlo simulation.

The further structure of the paper is as follows: *critical path problem* (CPP) and its mathematical formulation are described, the concept of the robust programming is introduced. We discuss the possibilities of using standard robust approach for the CPP. Afterwards, we show that the standard formulation is not fully suitable for the purposes of CPP and introduce an alternative modelling approach that helps to cover all robustness issues regarding the CPP. We demonstrate the proposed methodology on a small example and discuss the possible use of our approach in the end.

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, hlavaty@pef.czu.cz
[2] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, brozova@pef.czu.cz

2 Materials and methods

In this section, the formulation of the CPP is briefly introduced as well as one of the possible representations of this problem as MILP problem. Afterwards, it is shown how the aspects of uncertainty can be embedded into a generic MILP problem.

2.1 Critical path problem

In project time analysis, the basic question is what is the total project duration going to be and which of the project activities are critical. A change in duration of a critical activity always influences the total project duration whereas a change in duration of non-critical activity may or may not change the duration of the entire project depending on the slack it possesses. In any case, one needs to compute the total duration of the project and identify the critical activities first, then, it is possible to further analyse the possible changes and their influence on the current solution.

One can take advantage of the graph theory to depict the project as a set of consecutive or parallel activities. We will use the *activity on arc* approach where an edge of the graph represents an activity and its duration. A project can be modelled as a graph $G(V,E)$ that is directed, weighted and acyclic to comply with the logical structure of the project. Finding a critical path between the initial vertex v_0 and final vertex v_n and $n = |V|$ is the order of the graph G. Finding a critical path p_i between v_0 and v_n corresponds to *longest path problem* meaning that given a function $f: E \to \mathbb{R}$ and a directed acyclic graph G, we seek the maximum path $P = (v_0, \dots, v_n)$ that over all possible n maximizes $\sum_{i=1}^{n-1} f(e_{i,i+1})$.

Since our goal is to "robustify" the CPP we need to reformulate the stated graph problem as a mathematical programming problem. The *longest path problem* can be formulated as a MILP problem in the following way:

$$\begin{aligned} &\max \sum_{i=0}^{n-1}\sum_{j=1}^{n} t_{ij}x_{ij} \\ &\text{s.t.} \\ &\sum_{j\in Q_0} x_{0j} = 1 \\ &\sum_{k\in Q_j} x_{jk} = \sum_{i\in \mathcal{P}_j} x_{ij} \\ &-\sum_{i\in \mathcal{P}_n} x_{in} = -1 \\ &x_{ij} \in \{0,1\}, \forall i,j \end{aligned} \tag{1}$$

where $t_{ij} \geq 0$ is a weight of edge x_{ij} connecting vertices v_i and v_j (representing project activity i-j). The variable x_{ij} is binary such that $x_{ij} = 0$ if activity i-j is not a part of the longest path P and vice versa. The remaining set of constraints describe the structure of the graph in terms of directed edges entering a vertex v_j or leaving a vertex v_i. $\mathcal{P}_j$ is a set of edge indices entering a vertex v_j and Q_j is a set of edge indices leaving a vertex v_j.

2.2 Robust programming introduction

In the following paragraph, it is described how a MILP model can be transformed into its robust counterpart, that is, an extended MILP that seeks a robust optimal solution. Such solution is "resistant" against any deviations that might occur anywhere in the original deterministic model. For the description we utilize the approach of Γ robustness of Bertsimas and Sim [2,3]. We assume a generic MILP optimization model (2):

$$\begin{aligned} &\max \sum_{j=1}^{n} c_j x_j \\ &\text{s.t.} \\ &\sum_{j=1}^{n} a_{ij}x_j \leq b_i, i = 1, \dots, m \\ &x_j \in \mathbb{Z}, \forall j \in \mathcal{C}, \mathcal{C} \subseteq \{1, \dots, n\} \end{aligned} \tag{2}$$

$$\begin{aligned} &\max \sum_{j=1}^{n} (c_j \pm \delta_j^{\mathcal{C}}) x_j \\ &\text{s.t.} \\ &\sum_{j=1}^{n} a_{ij}\, x_j \leq b_i, i = 1, \dots, m \\ &x_j \in \mathbb{Z}, \forall j \in \mathcal{C}, \mathcal{C} \subseteq \{1, \dots, n\} \end{aligned} \tag{3}$$

When the presence of uncertainty is considered, we usually assume that the problem coefficients a_{ij}, b_i, c_j (or at least some of them) are not precisely defined. For the purposes of this article, let us consider the uncertainty to be present only within the set of coefficients c_j from now on. This leads to a new problem (3).

The (3) represents the reformulation of (2) with uncertain coefficients c_j. The uncertainty is expressed using deviations for any coefficient if need be. We assume any deviation δ_j^c to be any real nonzero positive number. It was illustrated by Ben-Tal, El Ghaoui and Nemirovski [1] that even a slight change in the original coefficient value may affect the optimal solution adversely. If there is a feasible realization of the program (2), then (3) is also feasible as the deviations in objective coefficients do not influence feasibility of the model. In the following section, we let ourselves inspired by approach of Bertsimas and Sim [2] and their concept of Γ robustness. For the robustness purposes, we introduce a parameter Γ, $0 \leq \Gamma \leq |\mathcal{U}|$ where $\mathcal{U} = \{j | \delta_j^c > 0\}$ describes the set of coefficients that are assumed to deviate from its expected value. The parameter Γ serves for the adjustment of robustness level for a given objective function. If it attains zero value then no robustness is considered. If it attains the value $|\mathcal{U}|$ then it is assumed that at most $|\mathcal{U}|$ coefficients in the objective are allowed to deviate from its deterministic value. Introducing Γ then leads to a new reformulation (4) of (2).

$$\max \sum_{j=1}^{n} c_j x_j - \max_{\{\mathcal{U} | \Gamma \leq |\mathcal{U}|\}} \left\{ \sum_{j \in \mathcal{U}} \delta_j^c x_j \right\}$$
$$\text{s.t.}$$
$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \,, i = 1, \dots, m \qquad (4)$$
$$x_j \in \mathbb{Z} \,, \forall j \in \mathcal{C} \,, \mathcal{C} \subseteq \{1, \dots, n\}$$

$$\max \Phi$$
$$\text{s.t.}$$
$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \,, i = 1, \dots, m$$
$$-\sum_{j=1}^{n} c_j x_j + \Gamma z + \sum_{j \in \mathcal{U}} u_j + \Phi \leq 0 \qquad (5)$$
$$z + u_j \geq \delta_j^c x_j \,, \forall j \in \mathcal{U}$$
$$z \geq 0$$
$$u_j \geq 0 \,, \forall j \in \mathcal{U}$$
$$x_j \in \mathbb{Z} \,, \forall j \in \mathcal{C} \,, \mathcal{C} \subseteq \{1, \dots, n\}$$

In (4), we want to find such solution that maximizes the original objective while seeking to maximize the set of deviations that will cause the maximum possible decrease i.e. maximum possible worsening of the objective. These leads to a max-min optimization problem that is generally difficult. Bertsimas and Sim [2] propose an approach how to overcome this obstacle by taking advantage of the dual model of (4) and embedding its parts into primal. Then the robust counterpart of (2) is defined as the program (5) where the parameter $\Gamma, 0 \leq \Gamma \leq |\mathcal{U}|$ controls the protection against uncertainty in the constraint, u_j is an auxiliary variable for each c_j that is considered uncertain, z is another auxiliary variable originating from merging the primal and dual of (4) and $\mathcal{U}$ indicates a set of indices j of those c_j for which the deviation is actually considered. The objective function was bounded with a new variable Φ. Note that (5) differs from the original formulation of Bertsimas and Sim [2] which is more general. As we lay some specific assumptions on the problem, we transferred the objective into the constraint in the same fashion that we earlier proposed in [8] and [9].

3 Results

The following chapter describes the robust formulation of CPP using the Γ robustness approach of Bertsimas and Sim [2]. It turns out that the simple reformulation of the problem according to their methodology does not provide satisfactory results in terms of practical use and therefore, we present an additional formulation that would help to cover all expected outcomes of the robust modelling of CPP.

3.1 Classical robust formulation of CPP

Finding a critical path and identifying critical activities in a project corresponds to finding a longest path in a network graph. This problem can be formulated as mathematical program as shown in (1). Using the robust model (5), the robust counterpart of the *longest path problem* is defined as follows:

$$\max \Phi$$
$$\text{s.t.}$$
$$-\sum_{i=0}^{n-1} \sum_{j=1}^{n} t_{ij} x_{ij} + \Gamma z + \sum_{(i,j) \in \mathcal{U}} u_{ij} + \Phi \leq 0 \qquad (6)$$
$$z + u_{ij} \geq \delta_{ij}^t x_{ij} \,, \forall (i,j) \in \mathcal{U}$$
$$\sum_{j \in Q_0} x_{0j} = 1$$

$$\sum_{k \in \mathcal{Q}_j} x_{jk} = \sum_{i \in \mathcal{P}_j} x_{ij}$$
$$-\sum_{i \in \mathcal{P}_n} x_{in} = -1$$
$$z \geq 0$$
$$u_{ij} \geq 0\,, \forall (i,j) \in \mathcal{U}$$
$$x_{ij} \in \{0,1\}\,, \forall i,j$$

The formulation (6) contains the original set of constraint from (1), the objective is replaced by a new variable Φ that allows to model the original objective as a constraint. Γ controls the maximum number of coefficients t_{ij} that are allowed to deviate from its deterministic value by δ_{ij}^t. Computing (6) yields the robust-optimal solution of (1) which can be understood as guaranteed objective in case of worsening of Γ unspecified coefficients t_{ij}. The worsening in this case means decreasing the value of objective as we seek to maximize it. Naturally, the robust-optimal solution will seek the guaranteed minimal longest path and identify those coefficients that would be a cause of it.

In the context of CPP, these results do not provide the protection against uncertainty as it is understood in the sense of Γ-robustness concept. The CPP indeed requires one to find a longest path in a graph which represents the critical path as such. However, in terms of project planning, the longest path cost is also understood as a minimum time needed to complete all activities and thus the entire project. If some of the edge weights i.e. activity durations should deviate from its value, the decision maker would understand the worsening in the sense of prolonging of the project time. But as the model (6) is maximizing, the worst-case scenarios are naturally those where the activity durations would shorten and thus possible worsen the objective.

This formulation is not useless, though. The model (6) finds such scenario where a specified number Γ of activities will shorten by a specific deviation. This scenario is then robust-optimal in that sense that it guarantees the minimum time needed to complete the project should some of its activities got shorter.

Of course, it would be more interesting for a decision maker to know what are the risks when a specified number Γ of activities would actually get prolonged and what is the worst-case scenario in that case. Unfortunately, it is not possible to use the classical concept of Γ-robustness for this purpose because that means we seek such solution that will improve the maximization function as much as possible considering there is a number Γ of deviations from the expected values.

3.2 Converse robust formulation of CPP

Because of the specific nature of CPP it is not possible to find the worst-case scenarios by simply applying the Γ-robustness approach on its MILP formulation. Instead, we propose our own formulation that would allow us to seek the worst-case scenarios in the same sense i.e. what is the worst possible prolongation of the project if we expect at most Γ coefficients to deviate from its expected value by a specified deviation. Now instead of (6), we propose a new formulation of robust counterpart to CPP:

$$\max \left(\sum_{i=0}^{n-1} \sum_{j=1}^{n} t_{ij} x_{ij} + \sum_{(i,j) \in \mathcal{U}} \delta_{ij}^t y_{ij} \right)$$

s.t.

$$y_{ij} \leq x_{ij}, \forall (i,j) \in \mathcal{U}$$
$$\sum_{(i,j) \in \mathcal{U}} y_{ij} \leq \Gamma$$
$$\sum_{j \in \mathcal{Q}_0} x_{0j} = 1$$
$$\sum_{k \in \mathcal{Q}_j} x_{jk} = \sum_{i \in \mathcal{P}_j} x_{ij}$$
$$-\sum_{i \in \mathcal{P}_n} x_{in} = -1$$
$$x_{ij} \in \{0,1\}\,, \forall i,j$$
$$y_{ij} \geq 0 \; \forall i,j$$

(7)

and let (7) be called *converse robust formulation* of CPP. The objective maximizes the longest path in the graph while it maximizes the deviations from expected activity durations. The new variable y_{ij} determines whether an individual deviation δ_{ij}^t will be counted into the robust-optimal solution or not. That requires y_{ij} to be binary. The binary character of y_{ij} is not defined in the constraints, yet it must be binary due to its relation to x_{ij} in the first constraint. The first constraint describes the logical relationship between variables x_{ij} and y_{ij}. If x_{ij} does not lie on the critical path ($x_{ij} = 0$) then necessarily $y_{ij} = 0$. If x_{ij} lies on the critical path ($x_{ij} = 1$) then $y_{ij} = 0$ or $y_{ij} = 1$ prioritizing those y_{ij} with higher δ_{ij}^t because we maximize. This comes from the fact that the critical activities x_{ij} influence the total project duration. Since y_{ij} is necessarily binary because of its relationship to binary x_{ij} then it is assured in the second constraint that at most Γ deviations will occur in the model. This formulation allows finding a (converse) robust-optimal solution that will choose into solution at most those Γ deviations that would prolong the total project duration the most.

3.3 Practical example

Let us demonstrate the described methodology on a small illustrative example. We assume a network graph $G(V, E), |V| = 5, |E| = 6$ as depicted in *figure 1a* and the alike graph in *figure 1b* where the deterministic durations are replaced by arbitrarily chosen intervals such that $t_{ij} \in \left[\, t_{ij} - \delta_{ij}^t \,,\; t_{ij} + \delta_{ij}^t \right]$. Here, it is considered w.l.o.g. that these intervals are symmetrical around its midpoint for the sake of clarity in results and avoiding overload of notation.

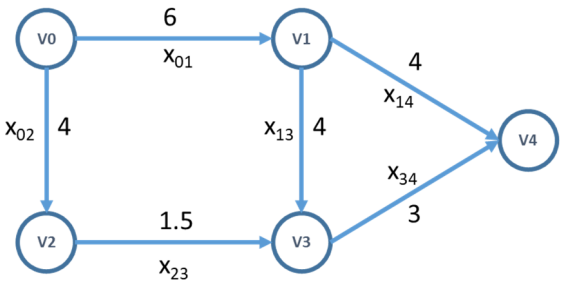


Figure 1a Graph with deterministic durations

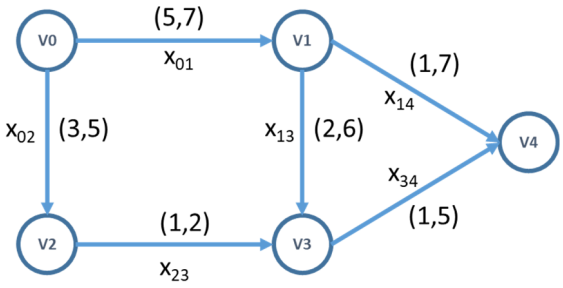


Figure 1b Graph with uncertain durations

The critical path in the deterministic version (*Figure 1a*) is given by the set of nodes $\{v_0, v_1, v_3, v_4\}$ and its cost is 13. Now, let us focus on the latter situation (*Figure 1b*) and discuss the two extreme situations should a Γ of unspecified coefficients change. These extreme scenarios under given conditions are finding the minimum time needed to finish the project if the durations get shorter (situation 3.1) or finding the minimum time needed to finish the project if the durations get longer (situation 3.2).

Starting with the situation 3.1, let us assume that a Γ of unspecified coefficients will reach the lower bound $t_{ij} - \delta_{ij}^t$ of the interval. If Γ is set to zero, then applying (6) will yield the same critical path and cost as in the deterministic case. Let $\Gamma = 1$ and recalculate. The resulting critical path remains $\{v_0, v_1, v_3, v_4\}$ and its cost decreased to 11. The critical path with the given cost indicates that either x_{13} decreased to 2 or x_{34} decreased to 1. It is easy to confirm that any other decrease of duration will not cause shorter critical path than this one. Pushing Γ further to higher values would eventually cause further shortening of the critical path down to the value of 8. This is the shortest critical path when all durations would shorten themselves to $t_{ij} - \delta_{ij}^t$. The solution for $\Gamma = 0$ and $\Gamma = 6$ is trivial (no change / all changes), however, it is hard to determine other solutions for $0 < \Gamma < 6$ where a combinatorial approach would have to be used as a brute force solution. For instance, if $\Gamma = 3$, one would have to try $C(6,3) = 20$ situations to find that the lowest possible critical path cost is 8.

With the opposite approach trying to find the worst case of the critical path (the longest one) when a number Γ of durations will prolong according to (7), let us use the same graph (*Figure 1a*) and setting durations to $t_{ij} + \delta_{ij}^t$. Setting $\Gamma = 0$ yields once again the same scenario as the deterministic case (1) where the cost equals 13 and the critical path is $\{v_0, v_1, v_3, v_4\}$. Setting $\Gamma = 1$ yields the new cost equal to 15. This is described as "what would be the worst (longest) critical path if one of the durations would prolong by δ_{ij}^t". By trying all six options, a reader can confirm that this will happen when x_{34} prolongs to 5. The following table (*Table 1*) shows the behavior of the critical path depending on the values of Γ.

	$\Gamma = 0$	$\Gamma = 1$	$\Gamma = 2$	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$
Critical path $\left(t_{ij} - \delta_{ij}^{t} \right)$	13	11	9	8	8	8	8
Critical path $\left(t_{ij} + \delta_{ij}^{t} \right)$	13	15	17	18	18	18	18

Table 1 Critical path depending on Γ

4 Conclusion

We addressed the *critical path problem* in terms of uncertain input in activity durations. We demonstrated how to treat uncertainties in durations without a need of probabilistic or fuzzy approach. First, it was shown how one can apply the concept of Γ robustness and it turned out that this approach is only applicable for finding the shortest guaranteed critical path in a project if the durations get shorter. We proposed our own methodology that allows finding the worst-case critical path when the durations get prolonged. Unlike [5], we focus on both situations (shortening/prolongation). If one is able to say that a percentage $\frac{\Gamma.100}{|\mathcal{U}|}$ of activities will get shorter or longer, the proposed algorithms are capable of finding best-case and worst-case scenarios for such a situation. Moreover in larger projects, if both algorithms are applied for every $\Gamma \in \mathbb{Z}, 0 \leq \Gamma \leq |\mathcal{U}|$, it will help to identify those activity durations that will frequently deviate from its expected value in the tested scenarios. This can be achieved without stochastic-based simulation like in [14]. Such activities then bear somewhat higher critical potential than other activities and should be focused on when the project is planned. A decision maker can be aware of these specific activities in advance and possibly secure them with more resources assigned.

Acknowledgment

This paper was supported by Czech Science Foundation (GAČR) project No. P403-18-04735S.

References

[1] Ben-Tal, A., El Ghaoui A.L. & Nemirovski A. (2009). *Robust optimization*. Princeton University Press.
[2] Bertsimas, D. & Sim, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming, 98*, 49–71.
[3] Bertsimas, D. & Sim, M. (2004). The price of robustness. *Operations Research, 52*, 35–53.
[4] Böhle, F., Heidling, E. & Schoper Y. (2016). A new orientation to deal with uncertainty in projects. *International Journal of Project Management, 34*, 1384–1392.
[5] Bruni, M.E. et al. (2017). An adjustable robust optimization model for the resource-constrained project scheduling problem with uncertain activity durations. *Omega, 71*, 66–84.
[6] Chakrabortty, R.K., Sarker, R.A. & Essam, D.L. (2017). Resource constrained project scheduling with uncertain activity durations. *Computers & Industrial Engineering, 112*, 537–550.
[7] Herroelen, W. & Leus, R. (2005). Project scheduling under uncertainty: survey and research potentials. *European Journal of Operational Research*, *165*, 289–306.
[8] Hlavatý, R. & Brožová, H. (2017). Robust optimization approach in transportation problem. In *35th Conference Mathematical Methods in Economics 2017* (pp. 225–230).
[9] Hlavatý, R. & Brožová, H. (2018). Matrix games with uncertain entries: A robust approach. In *36th Conference Mathematical Methods in Economics 2018* (pp. 139–144).
[10] Huemann, M. & Martinsuo, M. (2016). In project management, uncertainty is a great opportunity. *International Journal of Project Management, 34*, 1026–1027.
[11] Ke, H., Wang, L. & Huang, H. (2015). An uncertain model for RCPSP with solution robustness focusing on logistics project schedule. *International Journal of e-Navigation and Maritime Economy, 3*, 71–83.
[12] Kelly, J. & Walker, M. (1959). Critical-path planning and scheduling. In *Proceedings of the Eastern Joint Computer Conference, 1959* (pp. 160–173).
[13] Lechler, T.G., Gao, T. & Edington, B. (2013). *The Silver Lining of Project Uncertainties*. Project Management Institute.
[14] Na, W., Wuliang, P. & Hua, G. (2014). A robustness simulation method of project schedule based on the Monte Carlo method. *The Open Cybernetics & Systemics Journal, 8*, 254–258.
[15] Sadjadi, J.S., Pourmoayed, R. & Aryanezhad, M.B. (2012). A robust critical path in an environment with hybrid uncertainty. *Applied Soft Computing, 12*, 1087–1100.

Applications of Mathematical Optimization Approaches to Portfolio

Anlan Wang[1]

Abstract. In this paper the portfolio optimization problem is solved by applying mathematical optimization approaches. We apply the naive strategy to obtain a portfolio with equal weights. The efficient portfolios are also obtained by considering the standard deviation and the mean absolute deviation as the risk measure separately.
In our empirical analysis, based on the in-sample data, we construct the efficient frontiers by applying the Mean Variance approach and the Mean Absolute Deviation approach, and then we make the back-test of these efficient portfolios in the out-of-sample period to verify whether the strategies obtained by the optimization approaches work efficiently. In the back-test, the main performance measure of the portfolio is the Maximum Drawdown. To make the verification conclusive, we also generate random-weights portfolios and make hypothesis tests. By comparing the results, we conclude that for both of Mean Variance approach and Mean Absolute Deviation approach, they each obtains portfolios with efficiency in our empirical analysis.

Keywords: portfolio optimization, naive approach, Markowitz model, mean absolute deviation, maximum drawdown, random-weights portfolios, hypothesis tests.

JEL Classification: C12, G11
AMS Classification: 62F03**,** 46N10

1 Introduction

In the security trading markets, due to the influence of various uncertain factors, it's impossible to achieve the accurate prediction of the return of an investment, which makes the market always full of risks and opportunities. Therefore, on the premise of liquidity and security of investment funds, the most common investment strategy is to invest in a portfolio consisted of different securities to spread the risk. To maximize the profit of a portfolio investment, portfolio optimization is used to determine the best combination of securities with corresponding proportions.

The goal of this paper is to apply mathematical optimization approaches to portfolio, and then to verify the efficiency of optimized portfolios under the applied approaches.

Three portfolio optimization approaches are applied in this paper. The first one is a straightforward method called the naive approach, which is applied to invest the assets of a portfolio at a same weight. The second one is the Mean-Variance approach which is based on the framework of Modern Portfolio Theory. The third one is the Mean Absolute Deviation approach, it is proposed as an alternative to the classical Mean-Variance model.

The whole paper is divided into 5 sections. In section 1, we introduce the structure of the paper. In section 2, we make a literature review of the pioneers' researches. The theoretical basis of this paper is introduced in section 3. To verify the efficiency of the approaches applied in this paper, we make the empirical analysis in section 4. In section 5, we conclude the whole paper.

2 Literature Review

In 1952, Markowitz proposed a revolutionary approach for portfolio investments, which is called the Mean Variance (henceforth MV) model. In the Modern Portfolio Theory (henceforth MPT), it states that a portfolio investment's return and risk should not be viewed alone, moreover, a higher degree of risk means a higher potential return (Markowitz, 1952). Since then, with the development of algorithmic enhancements, this model has been improved a lot by considering real-life conditions. Based on this key thought of MPT, additional new constraints have been developed to the early Markowitz model by Konno and Yamazaki (1991) and Young (1998). Kolm, Tutuncu, and Fabozzi (2014) quantitatively focused on the portfolio investment by utilizing the trade-off between return and risk of securities. More than the single-objective MV model which aims to minimize risk when the level

[1] VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, anlan.wang.st@vsb.cz.

of return is fixed, Masmoudi and Abdelaziz (2018) focused on deterministic and stochastic multi-objective models to the portfolio optimization problems.

In the applications of portfolio optimization approaches, the mean absolute deviation (hence forth MAD) is also used as a risk measure instead of the standard deviation in some researches. The MAD approach was proposed as an alternative to the classical Markowitz model by Konno and Yamazaki (1992). The same as the statement of MPT, MAD method aims to find efficient portfolios that minimize the risk and maximize the overall expected return. In 2005, Mansini and Speranza (2005) proposed extensions for Konno's and Yamazaki's model to consider the incorporation of direct and indirect transaction costs of efficient portfolios, in addition, they also weighed the expected return and the risk by using a mixed-integer linear programming model.

Comparing with the Markowitz model, the MAD approach does not assume the normality of returns of securities, and it requires no computation or inversion of a covariance matrix, what's more, it solves a linear optimization rather than the quadratic optimization in Markowitz model. In this paper, the MV approach and simple MAD approach are both applied based on a same dataset in the empirical analysis, and the performances of strategies under these two approaches are also measured and compared.

3 Portfolio Optimization Approaches

In this section, three approaches of portfolio optimization problem are described with details. As a tool of the verification the efficiency of the strategies under these approaches, the hypothesis test is introduced. The descriptions in this section are made based on formulas.

3.1 Naive Approach

Comparing with other approaches in portfolio optimization problems, the most straightforward solution is called the naive approach, because when it is applied in portfolios investments, we simply invest the assets at a same weight $1/N$ (where N is the number of total assets) in a specified period, that's why naive approach is also named as $1/N$ approach.

The strategy obtained by naive approach is easy to implement because it does not rely neither on estimation of the asset returns nor on optimization. And assets allocation error caused by using the $1/N$ weights can turn out to be smaller than the error caused by using the weights from an optimizing model. Because the effect of estimation error on return probability distribution is so large, even the models designed explicitly to reduce the effect of estimation error achieve only modest success (Demiguel et al., 2009).

3.2 Markowitz Model

Markowitz model is based on the framework of analyzing the inter-relationships of return and risk in the portfolio optimization problems. In the Markowitz model, we denote x_i as the weight of asset i in a portfolio investment, and in our case, we exclude short sales, so the values of x_i satisfy $x_i \geq 0$ for all assets. We suppose that the expected stock return is identical to the average of the real returns within historical period (Zmeškal et al., 2004). If we denote $E(R_i)$ as the expected return of asset i in the sample period, then we can calculate the expected return of a portfolio $E(R_p)$ as follow,

$$E(R_p) = \sum_{i=1}^{N} x_i \cdot E(R_i) = x^T \cdot E(R) \tag{1}$$

where $x = [x_1, x_2, \dots, x_N]^T$, $E(R) = [E(R_1), E(R_2), \dots, E(R_N)]^T$, the sum of x_i of a portfolio equals to 1, and the expected return of portfolio $E(R_p)$ is the weighted average of $E(R_i)$.

As we mentioned, Markowitz model regards portfolio's variance or standard deviation as the risk measure, and they are calculated by the covariance $\sigma_{i,j}$ of the component assets for all asset pairs (i, j), we denote a $N \times N$ matrix as Q, and $Q = [\sigma_{i,j}, i = 1, 2, \dots, N, j = 1, 2, \dots, N]$, we show the calculations of variance ${\sigma_p}^2$ and standard deviation σ_p of a portfolio separately in equation (2) and equation (3),

$${\sigma_p}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \cdot \sigma_{i,j} \cdot x_j = x^T \cdot Q \cdot x \tag{2}$$

$$\sigma_p = \sqrt{\sigma_p{}^2} \tag{3}$$

where the standard deviation is the square root of variance.

From the MPT, we know that if an investor requires higher expected return, it means he or she must take on higher risk. In Markowitz model, based on the assumption of risk aversion, if investors meet the situations when the risk levels are different while the value of expected return is fixed, investors prefer to choose the portfolio with the minimum risk. In order to achieve this goal, the efficient frontier is applied.

The efficient frontier is the set of efficient portfolios that shows the highest expected return for a given level of risk or the lowest risk for a given level of expected return, to satisfy these conditions, we can construct an efficient frontier of Markowitz model based on the mathematical formulation (4).

$$\begin{aligned} &\text{minimize } \sigma_p{}^2 \\ &\text{subject to} \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, i = 1, \dots, N \end{aligned} \tag{4}$$

3.3 MAD Approach

The MAD approach was proposed as an alternative to the Markowitz model, the only difference in MAD approach is, changing the risk measure from the standard deviation into the mean absolute deviation of the portfolio's returns. The efficient frontier constructed by MAD approach can be constructed as following formulation,

$$\begin{aligned} &\text{minimize } MAD \\ &\text{subject to} \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, i = 1, \dots, N \end{aligned} \tag{5}$$

where the calculation of MAD is shown in equation (6),

$$MAD = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} x_i \cdot \left| R_{i,t} - E(R_i) \right|}{T} \tag{6}$$

where T is number of observations, $R_{i,t}$ is the return of asset i for each time t.

3.4 Hypothesis Tests

To evaluate the performance of the strategy portfolios under the applied approaches, we calculate the final value of wealth evolution, the maximum drawdown (henceforth MDD) of wealth, Sharpe ratio (henceforth SR) and average annual return (henceforth AAR) of the strategy portfolios as the performance measures.

However, even though the applied approaches are committed to optimize the portfolios, as the goal of this paper states, we need to verify the efficiency of strategies under the optimization approaches. So, to make the verification conclusive, not only the performances of the strategy portfolios are evaluated independently, the random-weights portfolios are also generated in our case to make the hypothesis tests by comparing their performances with those of the strategy portfolios. And in our case, the performance measure in the hypothesis tests is the MDD. As we know, MDD indicates the maximum loss from a peak to a trough of an investment's wealth evolutions up to time T, so, the smaller the MDD, the better performance of the portfolio.

As it literally means, the weights of assets in each random-weights portfolio are generated randomly, in our case, we set up 50,000 random-weights portfolios, and in each portfolio, the sum of the x_i is equal to 1. We know that a hypothesis test relies on the method of indirect proof (Arson, 2007). That is, to prove the hypothesis that we would like to demonstrate as correct, we show that an opposing hypothesis is incorrect. In our case, the strategy

portfolios under the optimization approaches are more likely to be demonstrated as efficient, so, according to the rule of hypothesis test, we can make the null hypothesis and the alternative hypothesis as below:

$$\text{null hypothesis} - H_0: MDD_s = MDD_r,$$

$$\text{alternative hypothesis} - H_A: MDD_s < MDD_r.$$

where MDD_s is the maximum drawdown of wealth evolutions of strategy portfolio, MDD_r is the maximum drawdown of wealth evolutions of random-weights portfolio. In our hypothesis test, the p-value is the proportion of the random-weights portfolios which meet $MDD_s = MDD_r$. We set the significance level at 10%. Then, if p-value < 10%, we reject H_0, which means the performance of strategy portfolio is better than that of the random-weights portfolio, so, the strategy is efficient; if p-value $\geq$ 10%, we fail to reject H_0, which means the performance of strategy portfolio make no difference from that of the random-weights portfolio, so, the strategy is inefficient in this case.

4 Empirical Analysis

4.1 Input Data

The dataset of the analysis is the weekly closing prices of 92 component stocks from Nasdaq 100, the Nasdaq 100 is a stock market index made up of 103 equity securities issued by 100 of the largest non-financial companies listed on the Nasdaq, there are 11 component stocks which are not included in our analysis due to the incomplete data in the analysis period, they are stocks of Alphabet Inc. Class C, American Airlines Group, Broadcom Inc., Charter Communications, Inc., Facebook, Inc., JD.com, Kraft Heinz Co, PayPal Holdings, Inc., Tesla, Inc., Tesla, Inc. and Workday, Inc..

The time duration in our analysis is 10 years, which is from October 5, 2008 to September 30, 2018. There are 522 weekly observations in total, and we divide the data evenly into two parts, they are in-sample part and out-of-sample part, the in-sample period is from October 5, 2008 to September 29, 2013, the out-of-sample period is from October 6, 2013 to September 30, 2018. In the application of MV approach and MAD approach, we both obtain 10 efficient portfolios based on the in-sample data and then make the back-test of these portfolios to verify the efficiency based on the out-of-sample data. For the naive approach and the analysis of random-weights portfolios, we only use the out-of-sample data. We assume the initial wealth to be 1 dollar in all portfolio investments.

The historical price evolutions of Nasdaq 100 of the in-sample part and the out-of-sample part are shown in Figure 1. We can see the price of Nasdaq 100 keeps increasing in the whole sample period, there is neither fall nor crisis.

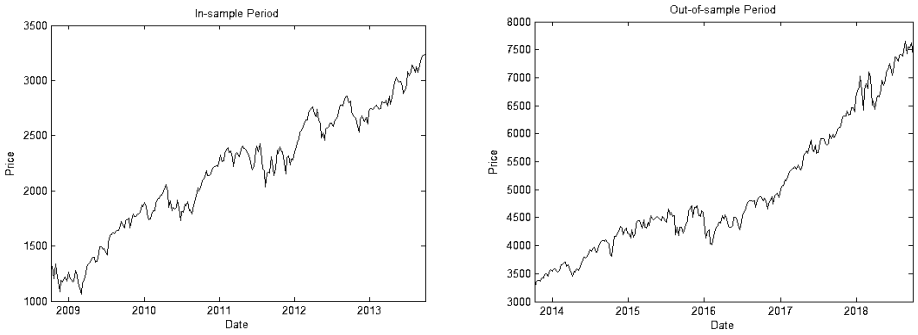


Figure 1 Historical price evolutions of Nasdaq 100

Source: https://www.investing.com/indices/nq-100-historical-data

4.2 Applications of Optimization Approaches to Portfolio

Naive Approach

We calculate the weekly returns of all the 92 stocks based on the out-of-sample data. As we mentioned before, in the naive approach, all the assets are invested at same weight $1/N$ (here $N = 92$). According to the returns of stocks and the weight, we calculate the weekly returns of the naive strategy portfolio according to equation (1). Next, based on the portfolio's weekly returns, we measure the performance of the naive strategy portfolio in the out-of-sample period, and the results are shown in Table 1. From Table 1, we find that the final wealth is $2.31,

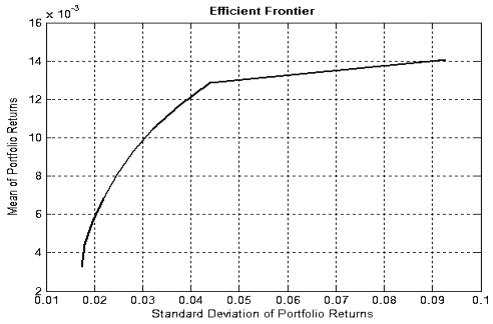
which has more than doubled the initial wealth, what's more, the SR is 14.36%, which means this portfolio is reliable.

Final wealth (dollar)	2.31
Mean of weekly portfolio returns	0.34%
Standard deviation of weekly portfolio returns	1.96%
AAR	19.39%
SR	14.36%
MDD	14.16%

Table 1 Performance of naive strategy portfolio

MV Approach vs. MAD Approach

MV approach and MAD approach are both applied in the in-sample period to construct the efficient frontiers. The efficient frontiers constructed by these two approaches are separately shown Figure 2 and Figure 3, and there are 10 efficient portfolios on each efficient frontier.



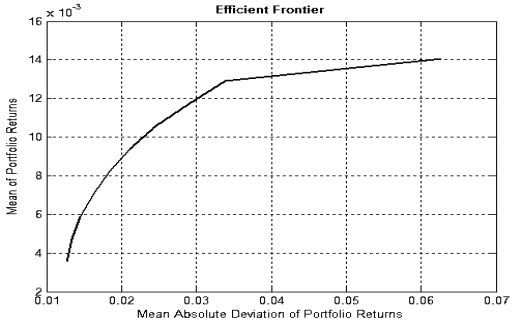


Figure 2 Efficient frontier of MV approach **Figure 3** Efficient frontier of MAD approach

From Figure 2, by applying the MV approach, we find the mean values of weekly portfolio returns on the efficient frontier are between 0.32% and 1.41%, and the standard deviations of weekly portfolio returns are between 1.74% and 9.29%. In Figure 3, by applying the MAD approach, we can see the mean values of weekly portfolio returns on the efficient frontier are between 0.35% and 1.41%, and the mean absolute deviations of weekly portfolio returns are between 1.27% and 6.28%.

In the out-of-sample period, we make the back-tests of the efficient portfolios which are obtained by MV approach and MAD approach separately, we show the performances of these efficient portfolios in the out-of-sample period in Table 2.

	Portfolio	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MV Approach	Mean of weekly *Rp*	0.20%	0.22%	0.25%	0.28%	0.30%	0.32%	0.33%	0.35%	0.37%	1.01%
	σ_p	1.73%	1.77%	1.88%	2.04%	2.22%	2.29%	2.34%	2.49%	2.82%	6.35%
	MDD	13.26%	12.15%	12.09%	12.99%	14.18%	14.06%	16.01%	19.72%	24.94%	36.77%
MAD Approach	Mean of weekly *Rp*	0.22%	0.24%	0.26%	0.29%	0.32%	0.33%	0.32%	0.34%	0.35%	1.01%
	σ_p	1.69%	1.74%	1.92%	2.08%	2.29%	2.27%	2.39%	2.59%	2.84%	6.35%
	MAD	2.00%	2.19%	2.47%	2.68%	2.86%	2.87%	2.89%	2.97%	3.01%	4.41%
	MDD	12.14%	12.42%	13.05%	15.12%	14.73%	14.11%	16.73%	21.56%	25.46%	36.77%

Table 2 Efficient portfolios' performances of out-of-sample period

From Table 2, we compare the performances of efficient portfolios under the two approaches. We find that when the values of mean of weekly portfolio returns under these two approaches are the same, on average, the value of standard deviation under MV approach is lower, and also, when the risk of portfolio is measured by MDD, given the same level of mean of weekly portfolio returns, the value of MDD under MV approach is lower.

4.3 Hypothesis Tests by Using MDD

We make the hypothesis tests to compare the performance of the strategy portfolios with that of random-weights portfolios. In the tests, we use the MDD as the performance measure, and according to the calculations, we know the critical value of the tests is 13.22%, and the p-values of the tests are shown in Table 3.

	Portfolio	1	2	3	4	5	6	7	8	9	10
p-value	Naive	34.87%	-	-	-	-	-	-	-	-	-
	MV	10.85%	0.96%	0.80%	6.62%	35.61%	31.38%	87.25%	99.92%	100.00%	100.00%
	MAD	0.92%	1.89%	7.40%	67.04%	54.54%	33.11%	94.61%	100.00%	100.00%	100.00%

Table 3 Hypothesis tests by using MDD

Based on the hypothesis assumed in section 3, we can see that under MV approach, for No.2, No.3 and No.4 portfolios, we reject H_0; and under MAD approach, for No. 1, No.2 and No. 3 portfolios, we reject H_0. That means these strategy portfolios perform better than the random-weights portfolios in the out-of-sample period.

5 Conclusion

The objective of this paper is, based on the specific dataset, we verify the efficiency of chosen strategies under the portfolio optimization approaches. Firstly, we apply the naive approach and measure the performance of the strategy portfolio. Then we make the back-tests of the efficient portfolios obtained by applying MV approach and MAD approach. In this sense, we find that all these obtained strategy portfolios earn profits in the out-of-sample period, which means these strategies are reliable. Secondly, the hypothesis tests are made in our case to verify whether the performance from strategy portfolio is better than that from random-weights portfolio. And in the tests, we use the maximum drawdown as the performance measure, which is used to indicate the maximum loss from a peak to a trough of an investment's wealth evolutions up to time T.

From the results of the hypothesis tests, we find that for both of MV approach and MAD approach, in our case, there exists three portfolios with efficiency for each one, which means if people invest in the 92 stocks followed by these efficient portfolios, there are high probabilities they can get profits with lower downside risk.

Acknowledgements

This work has been supported by the Czech Science Foundation (GACR) under project 17-19981S and SP2019/5, an SGS research project of VSB-TU Ostrava.

References

[1] Arson, D. (2007). *Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*. New York: Wiley.

[2] Demiguel, V., Garlappi, L. & Uppal, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies* (pp. 1915–1953).

[3] Kolm, P. N., Tutuncu, R. & Fabozzi, F. J. (2014). 60 Years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research* (pp. 356–371).

[4] Konno, H. & Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock-Market. *Management Science* (pp. 591–531).

[5] Kresta, A. (2015). *Financial Engineering in Matlab: Selected Approaches and Algorithms.* Ostrava: VSB-TU Ostrava.

[6] Mansini, R. & Speranza, M. G. (2005). An exact approach for portfolio selection with transaction costs and rounds. *IIE Transaction* (pp. 919–929).

[7] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* (pp. 77–91).

[8] Masmoudi, M. & Abdelaziz, F. B. (2018). Portfolio selection problem: A review of deterministic and stochastic multiple objective programming models. *Annals of Operations Research* (pp. 335–352).

[9] Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science* (pp. 673–683).

[10] Zmeškal, Z., Dluhošová, D. & Tichý, T. (2004). *Financial Models*. Ostrava: VŠB-TU Ostrava.

Sales Prediction Applying Linguistic Fuzzy Logic Forecaster

Aleš Kresta[1], Jan Kubečka[2], Tomáš Tichý[3]

Abstract. In this contribution, we focus on sales prediction by means of Linguistic Fuzzy Logic Forecaster (LFL-Forecaster). To be more specific, we compare the accuracy of the prediction obtained by means of this method and the prediction accuracy of standard approaches such as extrapolation of the time series and time series decomposition into trend and seasonal component. As the benchmark, we also apply the prediction based on the last known sales and average sales in the previous period. The LFL-forecaster combines the fuzzy transform technique to extract the trend part with fuzzy natural logic in order to forecast future values. Both methods apply the principles of fuzzy sets. From the obtained results, we demonstrate that for selected time series of sales the LFL-Forecaster provides the most accurate prediction in the out-of-sample period, however even this method is prone to the changes in the length of input data and structural breaks.

Keywords: forecasting, fuzzy sets, prediction, sales.

JEL Classification: C53, G30
AMS Classification: 60G25

1 Introduction

Obtaining reliable prediction of future sales is a key aspect of financial planning in every company. Sales typically depend on various endogenous variables, such as GDP, unemployment or CPI, which cannot be planned easily, or product-related parameters (e.g., life period), which in contrast can be well predicted. In addition to pure economic quantities, we can often find a good proxy of sales trends when variables specific to a given product are considered, such as natality for infant products or average temperature for leisure activities and related products (see e.g. [2]). Of course, such a relation should be supported by well-established theory and not represent a result of pure statistics alone (see, e.g., [1], for relevant discussion).

However, in some cases we are not able to find the variables, the sales depend on, such is the case of sales time series studied by Kubečka [4], or these variables cannot be predicted accurately. In this case, we want to predict the future sales as best as we can from the historical data. Generally, the most promising approach is to decompose the time series into components such as trend, cyclical and seasonal components. Generally, these components can be predicted more accurately than the original time series. This approach is also utilized by Linguistic Fuzzy Logic Forecaster (LFL-Forecaster[4]). In this contribution, we focus on comparison of sales prediction accuracy by applying this method and compare it to the more standard methods. We study the same times series as in [4].

We proceed as follows. In the next section, the forecasting of (financial) variables is discussed with special attention to the theoretical foundations of LFL-Forecaster. In section 3, we provide the definition of applied accuracy measures. In section 4, we display the results and section 5 concludes the paper.

2 Prediction Methods and Models

A sequence of random variables in time is referred to as a time series. Although in finance we usually assume continuous distributions, their observations are by nature discrete and thus a sequence in time is always finite. Assume we have a historical financial time series of sales $\{y_t\}$. Usually we split the whole time series into two parts: in sample part $\{y_t\}_{t=1}^{n}$ and out-of-sample part $\{y_t\}_{t=n+1}^{n+m}$, where n is the length of data utilized for parameter estimation and m is the length of data left for judging the accuracy of prediction.

In the next lines we briefly review a standard approaches for forecasting of time series; subsequently, an advanced approach based on fuzzy-linguistic definitions will be discussed.

[1] VSB-TU Ostrava, Department of Finance, Sokolska 33, 701 21, Ostrava, Czech Republic, ales.kresta@vsb.cz
[2] VSB-TU Ostrava, Department of Finance, Sokolska 33, 701 21, Ostrava, Czech Republic, jan.kubecka.st1@vsb.cz
[3] VSB-TU Ostrava, Department of Finance, Sokolska 33, 701 21, Ostrava, Czech Republic, tomas.tichy@vsb.cz
[4] http://irafm.osu.cz/en/c110_lfl-forecaster/

2.1 Prediction by Means of Historical Average

The simplest prediction of the future values can be based on the last observed value,

$$\hat{y}_{n+t} = y_n, t = 1,...,\infty, \tag{1}$$

or average of historical values,

$$\hat{y}_{n+t} = \frac{1}{k} \sum_{i=n-k+1}^{n} y_i, t = 1,...,\infty. \tag{2}$$

Such the prediction is very simple and will be assumed as a benchmark in our study.

2.2 Simple Extrapolation of the Time Series

The principle of this method is to find a suitable function that would describe the evolution of the selected time series values. In general, a curve can be found for a given time series that passes directly or at least very close to the values of this series. Then, by extending the values beyond the known values we find the forecasted future values. This method is generally very inaccurate and creates a higher risk of producing meaningless results.

The suitable function can be linear, quadratic, cubic, exponential or any other form of the function. The more parameters needed to be estimated, the closer the function values to the observed values in-sample, however, the higher the risk of overfitting. The example of cubic function is as follows,

$$\hat{y}_{n+t} = a \cdot (n+t)^3 + b \cdot (n+t)^2 + c \cdot (n+t) + d, t = 1,...,\infty, \tag{3}$$

where *a*, *b*, *c* and *d* are estimated parameters from in-sample data. In the paper, we apply different functions, however, we report only the results of the cubic function and functions with the lowest in-sample RMSE.

2.3 Time Series Decomposition

Some time series, and especially those of an economic nature, can be broken down into several specific components, namely trend, seasonal, cyclic and residual component. Decomposition of the time series into these components is performed with the expectation that the regular behavior of the time series can be better recognized in the individual components of the decomposed time series than in the the non-decomposed series.

In the simple form, the time series can be decomposed applying additive decomposition,

$$y_t = T_t + C_t + S_t + \varepsilon_t, \tag{4}$$

or multiplicative decomposition,

$$y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t, \tag{5}$$

where *T* represents the trend component, *C* represents the cyclical component, *S* is seasonality and ε is random noise, i.e. the error term. While the trend part should depict long-term change in the random variable, the seasonal component should capure short periodically repeating changes, such as time of year/week/day. On the other hand, random noise is a kind of error between the predicted value and real observation, which cannot be explained by any function of time (its mean should be zero). There are different approaches how to estimate particular components. In the paper we assume only trend component modelled by linear and exponential function and seasonality modelled by additive and multiplicative decompositions.

In the next subsection, we present more complex approach utilizing linguistic fuzzy logic in decomposition.

2.4 LFL-Forecaster

The LFL-forecaster combines the fuzzy transform (F-transform) technique to extract the trend part with fuzzy natural logic (FNL) in order to forecast future values, see [5]. Both methods apply the principles of fuzzy sets, i.e., a function $A: U \rightarrow [0,1]$ with *U* being the universe and $[0,1]$ being its support set on standard algebra, to time series analysis.

F-transform

The principles of F-transform were formulated by Perfilieva (see e.g. [9] as well as [3]). The key idea is to transform a continuous function $f: [a, b] \rightarrow R$ to a finite vector of numbers (*direct F-transform*) and then transform it back (*inverse F-transform*) to the original space so that an approximating function $\hat{f}$ is obtained. The procedure is as follows.

First, form a *fuzzy partition* of a given domain $[a, b]$, which consists of a finite set of fuzzy sets $A = \{A_0, \dots, A_n\}$, $n \geq 2$, fulfilling the axioms of *normality*, *locality*, *continuity*, *unimodality*, *and orthogonality*, defined over the nodes constructed on the domain, i.e., $a = a_0, \dots, a_n = b$. The membership functions of $A_0, \dots, A_n$ in the partition *A* are called *basic functions*; such partition is called *h-uniform* if the nodes $a_0, \dots a_n$ are *h-equidistant*, i.e., $a_{k+1} = a_k + h$ with $h = (b - a)/n$ for all $k = 0, \dots, n - 1$.

Having the partition *A* in place, we can define the direct F-transform as a vector $\mathbf{F}[f] = (F_0[f], \dots, F_n[f])$ with each *k*-th component $F_k[f]$ specified as follows:

$$F_k[f] = \frac{\int_a^b f(x)A_k(x)\mathrm{dx}}{\int_a^b A_k(x)\mathrm{dx}}, k = 0, \dots n. \tag{6}$$

Such a partition is called *h*-uniform if the nodes $a_0, \dots a_n$ are *h*-equidistant, i.e., $a_{k+1} = a_k + h$ with $h = (b - a)/n$ for all $k = 0, \dots, n - 1$.

Subsequently, the inverse F-transform, which helps us attain an approximation of the original function, is a continuous function $\hat{f}: [a, b] \rightarrow R$ such that

$$\hat{f}(x) = \sum_{k=0}^{n} F_k[f]A_k(x)\,, x \in [a, b]. \tag{7}$$

Obviously, $\hat{f}$ should converge to f with $n \rightarrow \infty$.

More details about the F-transform as well as all related proofs can be found in [9] and [10]. Note also that the definition above is in fact, F^0-transform, i.e., a zero-degree F-transform, because the components are real numbers. Clearly, the procedure can be easily generalised to F^m-transform, for example, with F^1-tranform the components are formed by polynomials, such as triangular or cosine functions.

Fuzzy natural logic

Following [6], fuzzy natural logic (FNL) is a group of mathematical theories that extends mathematical fuzzy logic in a narrow sense, thereby developing a mathematical model of special human reasoning when natural language is employed.

It is based on a set *EvExpr* of evaluative linguistic expressions. For our purposes, the most useful subclass of *EvExpr* is a class of *simple evaluative expressions* structured as:

$$\langle \text{linguistic hedge} \rangle \langle \text{TE adjective} \rangle. \tag{8}$$

Here, TE adjective forms fundamental evaluative trichotomy, with typical examples being *gradable adjectives* (such as *strong*, *long*, etc.) and *evaluative adjectives* (such as *low*, *medium*, *high*, etc.). On the other hand, the so-called linguistic hedge represents adverbial modifications, such as *intensifying adverbs* of *narrowing sense* (such as *extremely*, *significantly*, etc.) or *widening sense* (such as *more or less*, *roughly*, etc.). The idea behind its utilisation in time series analysis and forecasting is that FNL can allow us to transform past observations of the trend/cycle into the future, using natural language.

Trend estimation and its forecasting

Estimation of the trend/cycle part of the time series can be obtained following the theorem in [5] and [7]:

Theorem. *Let X_t be a realisation of random variable over interval $[0, b]$. Let us construct an h-uniform fuzzy partition A over nodes $a_0, \dots a_n$ with $h = d\bar{T}$, where $\bar{T} = 2\pi/\lambda$ for a minimal $\lambda = min\{\lambda_1, \dots, \lambda_r\}$ of seasonal frequencies with r observations and a real number $d \geq 1$. If we compute a direct F-transform $F[X]$, then for the corresponding inverse F-transform $\hat{X}$ of X there is a certain small number D converging to 0 for $d \rightarrow \infty$ such that*

$$\left|\hat{X}_t - T_t\right| \leq D, t = |a_1, a_{n-1}|. \tag{9}$$

As soon as the trend has been estimated, the new task is to extrapolate the known values of *X* (or rather $\hat{X}$) into *l* future intervals (a so-called forecasting) using the implications of fuzzy natural logic. Given that direct F-transform leads to a vector of components,

$$\mathbf{F}[X] = (F_1[X], \ldots, F_{n-1}[X]), \tag{10}$$

where each such component represent a weighted average of values of X_t in the area of width $2h$, they can be used as data to learn a linguistic description. Subsequently, using FNL, the future components can be forecasted,

$$F_n[X], \ldots, F_{n+l}[X]. \tag{11}$$

Finally, the inverse transform is used to calculate the trend/cycle development. For more details of possible linguistic descriptions, see, e.g., [8] and [5].

3 Applied Accuracy Measures

In order to examine the accuracy of the predictions we apply a standard accuracy measures, which generalize the differences between forecasted values ($\hat{y}$) and truly observed values (y) into one number. Based on the way how the errors are summarized we can define the root mean squared error (RMSE),

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=n+1}^{n+m}\left(y_t - \hat{y}_t\right)^2}, \tag{12}$$

which should represent some "mean error", however due to the squaring of the errors, which cases that the larger errors have progressively higher penalty, the interpretability is not easy. Some researchers have recommended the use of the mean absolute error (MAE),

$$MAE = \frac{1}{m}\sum_{t=n+1}^{n+m}\left|y_t - \hat{y}_t\right|, \tag{13}$$

which clearly represent the mean of the errors, i.e. the mean difference between the forecasted and truly observed values without considering the sign. Another standardly applied measure is the mean absolute percentage error (MAPE),

$$MAPE = \frac{1}{m}\sum_{t=n+1}^{n+m}\left|\frac{y_t - \hat{y}_t}{y_t}\right|, \tag{14}$$

which however puts a heavier penalty on negative errors ($y_t < \hat{y}_t$) than on positive errors ($y_t > \hat{y}_t$) and thus is biased in that it will systematically select a method whose forecasts are too low. The last accuracy measure we consider focus on the accuracy of the directions of further development. We can define the mean directional accuracy (MDA),

$$MDA = \frac{1}{m}\sum_{t=n+1}^{n+m} 1_{(y_t - y_{t-1})\cdot(\hat{y}_t - \hat{y}_{t-1})>0}, \tag{15}$$

In all the above measures n is the length of the data applied in model calibration and m represent the length of forecasted data.

4 Sales Forecasting

In this section we present the results of the sales prediction of a selected company obtained by means of the above described methodology. The selected company operates in the area of the purchase of secondary raw materials and precious metals (CZ-NACE 38.32). All methods are based on the same time series of sales as shown in Figure 1. From the figure, it can be seen that the sales do not have a clear trend in the given period. Furthermore, there is a clear difference between the average value of sales in 2010-2012 (approximately 103M CZK), 2013-2015 (approximately 32M CZK) and 2016-2017 (approximately 19M CZK). Also remarkable is the extreme value in the first month of 2018, when the value of sales reached almost 250M CZK. This value was due to the situation in the holding structure, concretely the selected subsidiary took over the responsibilities of the other subsidiaries.

Due to the clear structural breaks in time series, we assume three different intervals of in-sample period: long (8 years; 2010-2017), medium (5 years; 2013-2017) and short (3 years, 2015-2017) in order to compare the robustness of the methods for changes in input data (and structural breaks).

Data for year 2018 are left as out-of-sample data, i.e. these data do not enter in estimation of selected models. We present the accuracy measures for period from 2/2018 until 10/2018 as there is an extreme value in January,

however, we calculated also accuracy measures for the January-October period and obtained similar results. Kubečka [4] found that applying linear regression on the selected explanatory variables provides statistically insignificant parameters and inaccurate prediction of future values, thus, we do not consider this approach.

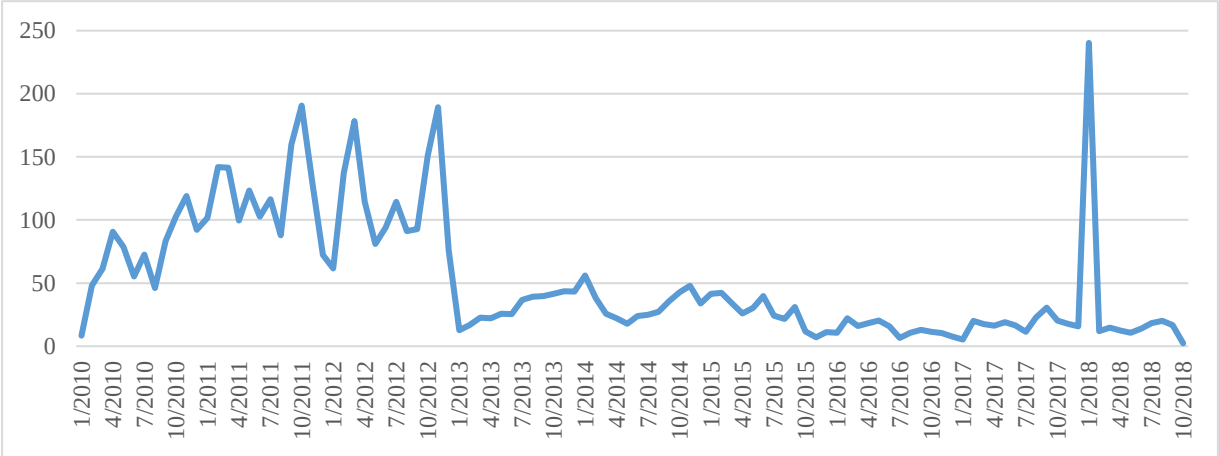


Figure 1 Historical sales (in millions CZK)

The results of the selected forecasting methods are depicted in Table 1. In the table, we present the accuracy measures of the out-of-sample forecasts in the following order. The first method represents the prediction based on the last observed in-sample sales, i.e. sales in December 2017. As it is obvious from the results, this benchmark method provides very good prediction, which even other standard methods are unable to achieve. To be concrete, the average difference between prediction and observation is approximately 4M CZK (approximately 5M CZK when measured by RMSE), however, on average the forecast is wrong by more than 82%.

The most of the inaccuracy of this and also other models is caused by the last month (October, 2018), in which the observed sales are 2.6M CZK (see Figure 1) while the model-forecasted sales are around 20M CZK. If we remove this month from the out-of-sample comparison, the measures improves a lot, however, we are not sure what happened in this month and whether it is a normal value (i.e. new structural break in the time series) or extreme value caused by external factors (such as January, 2018).

Prediction method	RMSE	MAE	MAPE	MDA
sales in 12/2017	5.09	3.99	82.61%	62.50%
average sales in 2013-2017 (medium)	11.33	10.90	168.97%	50.00%
average sales in 2015-2017 (short)	7.37	6.15	115.94%	75.00%
cubic extrapolation (long)	61.33	62.13	831.73%	50.00%
cubic extrapolation (medium)	8.43	7.47	50.30%	50.00%
cubic extrapolation (short)	24.76	24.60	359.12%	50.00%
exponential extrapolation (long)	5.54	4.77	57.30%	50.00%
linear extrapolation (medium)	4.99	4.16	56.91%	75.00%
inverse extrapolation (short)	5.79	5.08	56.84%	50.00%
multiplicative decomposition, linear trend (long)	19.88	20.30	197.10%	50.00%
multiplicative decomposition, linear trend (medium)	5.74	5.04	63.39%	62.50%
additive decomposition, linear trend (long)	23.89	22.72	176.68%	50.00%
additive decomposition, linear trend (medium)	6.18	5.47	70.99%	62.50%
multiplicative decomposition, exponential trend (long)	9.68	7.77	153.68%	50.00%
multiplicative decomposition, exponential trend (medium)	8.21	7.02	121.28%	50.00%
multiplicative decomposition, exponential trend (short)	17.43	16.33	198.72%	50.00%
LFL Forecaster (long)	24.24	24.69	274.58%	50.00%
LFL Forecaster (medium)	4.67	3.64	72.09%	75.00%

Table 1 Accuracy measures in out-of-sample period (in millions CZK and %)

Prediction of sales based on average value in medium and short in-sample period provides worse forecast than the sales in 12/2017. We can also observe that the medium period provides much worse prediction, which is caused by the fact that there was a structural break in time series in 2015 as we discussed above.

Next, we show the results of cubic extrapolation, which works best for medium-length in-sample period. Then we present the results of the extrapolation with lowest RMSE in-sample for long/medium/short period (concretely exponential, linear and inverse functions). For these methods, the results are mixed – RMSE and MAE are worse than in case of 12/2017 sales prediction, but MAPE is generally better.

Next group of presented results are those of time series decomposition (multiplicative and additive) with different trend predictions (linear and exponential). From the results, we can observe that these methods work best in the case of medium period (2013-2017). However, even in this period their accuracy is worse than accuracy of 12/2017 sales forecast, except for MAPE.

Finally, last two rows correspond to the prediction obtained by LFL Forecaster (for short period there is not enough data). The results are mixed. For long period the accuracy is worse than 12/2017 sales and other methods applied in long in-sample period. For medium-length period, the prediction has lowest average error (RMSE, MAE) and the best directional accuracy (MDA). Mean percentage error is better than of methods based on 12/2017 sales prediction and average sales in medium/long period.

5 Conclusion

Obtaining reliable prediction of future sales is a key aspect of financial planning in every company. In this contribution, we focus on comparison of sales prediction accuracy by applying LFL-Forecaster and compare it to more standard methods in case of time series, for which we cannot find the variables, on which the sales depend. From the obtained results, we demonstrate that for selected time series of sales the LFL-Forecaster provides the most accurate prediction in the out-of-sample period, however even this method is prone to the changes of input data length and structural breaks.

Acknowledgements

The first author was supported through the Czech Science Foundation (GACR) under project 18-13951S: The third author was supported through the ESF in "Science without borders" project, reg. no. CZ.02.2.69/0.0/0.0/16_027/0008463 within the Operational Programme Research, Development and Education and SP2019/5, an SGS research project of VSB-TU Ostrava. The support is appreciated.

References

[1] Armstrong, J.S., Green, K.C. & Graefed, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, 68(8), 1717–1731.
[2] Dvořáčková, H., Kresta, A. & Tichý, T. (2018). Sales Prediction in the Ice Category Applying Fuzzy Sets Theory. *Ekonomická revue*, 21(2), 35–41.
[3] Holčapek, M. & Tichý, T. (2011). A smoothing filter based on fuzzy transform. *Fuzzy Sets and Systems*, 180(1), 69–97.
[4] Kubečka, J. (2018). *Predikce finančních časových řad podniku*. Master Thesis, Ostrava: VSB-TU Ostrava.
[5] Novák, V. (2016). Linguistic characterization of time series. *Fuzzy Sets and Systems*, 285, 52–72.
[6] Novák, V. (1995). Towards formalized integrated theory of fuzzy logic. In: Z. Bien, K. Min (Eds.): *Fuzzy Logic and Its Applications to Engineering, Information Sciences, and Intelligent Systems* (pp. 353–363). Dordrecht: Kluwer.
[7] Novák, V., Perfilieva, I., Holčapek, M. & Kreinovich, V. (2014). Filtering out high frequencies in time series using F-transform. *Information Sciences*, 274, 192–209.
[8] Novák, V., Štěpnička, M., Dvořák, A., Perfilieva, I., Pavliska, V. & Vavříčková, L. (2010). Analysis of seasonal time series using fuzzy approach. *International Journal of General Systems*, 39, 305–328.
[9] Perfilieva, I. (2006). Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157, 993–1023.
[10] Perfilieva, I., Daňková, M. & Bede, B. (2011). Towards a higher degree F-transform. *Fuzzy Sets and Systems*, 180, 3–19.

Decision-making support as part of the sustainable investment in unit trust funds

Oldřich Trenz[1], Oldřich Faldík[2], Sylvie Formánková[3], Jan Kolomazník[4], František Ostřížek[5]

Abstract. The article deals with decision-making support in the field of sustainable investment in unit trust funds with an ESG evaluation, in accord with socially responsible investment (SRI). The individual methods, which take into account the analysis of long-term risks in conjunction with the expected investment return, will be analyzed and concretized. This will be complemented with the description of the present state in the field of sustainable investing. The applied methods will include the quantitative measurement of risks where we will analyze the dependence of profitability and sustainability in the form of a selective portfolio theory (Markowitz). Furthermore, the Data Envelopment Analysis (DEA) method will be implemented, and this in order to contrast the relationships between profitability, risk-management and sustainable development. Last but not least, a formalization of the relation of the risk involved and the investment returns will be done, with the aid of a select artificial intelligence method. The results will be used for the support of the decision-making process in the field of unit trust investments.

Keywords: sustainable investing, sustainability, SRI, decision making, funds.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Investments in financial markets can be very deceitful as it has become evident in the degree of the last financial crisis in 2008. Some financial opportunities may be too risky (technological start-ups), or bringing little profit (stocks and shares), or rather with little liquidity (collections of pieces of art). Many investors have begun to look for investment opportunities on financial markets and that with respect to a well-diversified portfolio. Together with the growth of small investors personal property are these investors considering investments in unit trusts with the emphasis on security in their old age. New investors may often not orient themselves well in investments opportunities, in the inexhaustible amount of investment funds. The main criterion for the selection of investment opportunities is the aim of the investment and the composition of the investment portfolio, in other words in the form of the connection between the risk and the investments profits. In this context we are talking about so called "age of responsibility" where this issue was even part of the inaugural speech of the US president Barack Obama (Washington Post, January 21, 2009).

During the last few years, investors have been more and more interested in those investments in which the rules of socially responsible investments (SRI) [11] are being applied that are in agreement with the companies' social responsibility (CSR) [20]. To assess those investments ESG evaluation criteria are very often applied (the environment, social responsibility and corporate governance) [8]. The principles of sustainability are introduced in many visions and approaches and this projects itself into the plans of different organizations, political parties and the support within individual countries; in the USA this idea is called the "Green New Deal" and is aiming at the support of the sustainable development and prevention of economic crisis [22]. In 2016, 106 billion USD was like this released from the total of 787 billion USD for this activity and for the form of socially responsible investment with the vision of an increase of this amount in the following years [2].

[1] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.trenz@mendelu.cz.
[2] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.faldik@mendelu.cz.
[3] Mendel University in Brno/Faculty of Business and Economics/Department of Management, Zemědělská 1, 613 00 Brno, Czech Republic, sylvie.formankova@mendelu.cz.
[4] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, jan.kolomaznik@mendelu.cz.
[5] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, frantisek.ostrizek@mendelu.cz.

When the investments into the area of mutual funds are taken into consideration in connection with CSR and SRI, we then talk about so called "responsible investment" [6], "ethical investment" or "green investments" [22]. Sustainable investments fall within the industrial areas oriented to renewable energy sources, sustainable transportation, sustainable agriculture and soil management, water management, waste management, education, health care and many other areas [1]. ESG mutual funds that select investment opportunities from these areas meet the high standards of the environment protection, ecology, ethics, work with the employees and the transparent management of the company [8].

Aspects of sustainable investment and an effort to influence the form of further development motivate even smaller investors to participate in investment opportunities in accordance with the rules of sustainable investment [3]. At present, in accordance with the support on the national and international levels (obligatory reporting), the position of accessible information about suitable investment opportunities has been facilitated [20]. Many organizations are providing information about individual segments (bonds, mutual funds, ownership interests, etc.), however they often do not provide common users with the tools for selection of a suitable portfolio product and its evaluation. The aim of the article is the analysis of selected methods for assessment of investment opportunities in mutual funds with ESG evaluation and formulation of opinions about the appropriateness of integration of the selected method into the software tool intended to support decision-making in small and medium-sized investors.

2 Material and Methods

The investor, while making a decision within the frame of possible investment opportunities, assesses the size of their investment into a given financial institution (fund, investment portfolio) and that in connection with the long-term risk and expected profit from this investment. To assess these aspects, it is possible to use selected methods assessing the risk dependence and profit dependence in the selected period.

A study performed in 2015 by the Oxford University and the Arabesque Partners analyzed over 200 sources the result being that 88 % of the reviewed sources attained the result of a positive SRI impact on the investments, in the form of their higher performance [5].

The funds' efficiency is always assessed for a given time period, for the case of this article the efficiencies of funds will be taken into account for the last 5 years, 3 years, 1 year and the last 6 months. The term of efficiency in this concept means the percentage growth/decline of the fund for the monitored period. The situation is more complex when assessing the rating (degree of risk). There are more methodologies assessing the quality of the given fund and the conversion between different assessment approaches is often difficult. For the needs of funds' data assessment, the SRRI methodology (a synthetic indicator of risk and profitability) will be used; one which is, for example, being utilized by the Erste Group company. The value of this indicator can reach an interval of 1–7 and relate the degree of risk and volatility. The threshold value equals small risk, the value of 3 medium risk and the value of 7 high risk [10].

2.1 Portfolio Theory – the Selective Markowitz Model

The Portfolio Theory – the Selective Markowitz Model puts into context the expected portfolio revenue, portfolio risk, the effective boundary of the investment into the assessed funds and allows us to create an optimal investment portfolio. The basic assumption of the model is that the investors think rationally, all of them invest for an equally long period.

When compiling a portfolio from *n* investment instruments in a way where particular weight w is assigned to each instrument so that the total weight equals 1. The individual weights in this concept can be even of negative value, which indicates a shortening of a given instrument. The expected profit is then given by the formula:

$$E(R_P) = \sum_{i=1}^{n} \omega_i . R_i \tag{1}$$

So the portfolio profit is the weighted average of profits of individual actives.

The Markowitz optimal portfolio is such a X_{opt} portfolio for which there exists an indifference curve such, that

$$\left\{\left(R_{opt}, \sigma_{opt}\right)\right\} = u_k \cap EM\ , \tag{2}$$

where a set of permissible portfolios is being assessed and where *EM* is a set of effective portfolios in the Markowitz sense [9]. The method is analyzed in more detail incl. limiting conditions in the first Markowitz's articles [15], [16]. For the actual calculation, the Python programming language environment will be used as well as the available source codes of the Markowitz method modified for this programming language [18].

2.2 Data Envelopment Analysis (DEA)

The method of data envelopment (DEA – Data Envelopment Analysis) is an approach utilized for the assessment of efficiency, or the performance of homogenous production units (DMU – Decision Making Unit). Each of these production units consumes inputs to create the required effects (outputs). This is an optimization method from the field of multi-criteria decision-making [4]. The aim of this method is to classify the production units as effective or non-effective by comparing the inputs and output sizes [7]. The approach enables us to compare a given sub-unit with the best units regarding the effectivity. The DEA uses a method of an estimate of the production function which is based on the theory of linear programming [13].

The ratio of the input and output is called the efficiency of production units. This ratio is a relation of:

$$efficiency = \frac{inputs}{outputs} \tag{3}$$

With the total efficiency of production units there may occur a situation where the whole set of inputs and outputs is given. In this case, the calculation of the relative degree of efficiency is used according to the relation:

$$\Phi_k = \frac{\sum_{j=1}^{n} u_j y_{jk}}{\sum_{i=1}^{m} v_i x_{ik}}, k = 1,..., p, \tag{4}$$

where v_i and u_j are the individual input and output weights for all assessed units (total of p units). Each unit for its production consumes m inputs and n outputs, x_{ik} is the inputs number consumed by the unit k, and y_{jk} is the number of output j produced by the unit k. The relation for the relative degree of efficiency can be generalized in the sense of different weights of inputs and outputs for the assessed unit, the weights are then determined so that the efficiency of the individual units is maximal. A unit is effective if the efficiency coefficient equals one. If the efficiency value of a given unit is lower, it means that there exists a production unit with a higher efficiency value [13]. For the efficiency calculation a script in the Maple environment will be used [12].

2.3 Self-organizing neural network (SOM)

Artificial neural networks are one of many computing models used in the field of artificial intelligence [19]. In application, multilayer neural networks are often deployed for classification, prediction or approximation tasks because assessed learning data are necessary here. This is the learning with the teacher type. Another type of neural networks are self-organizing neural networks that boast a principle similar to cluster methods [14]. The advantage of these networks is that the learning data do not have to be assessed (learning without a teacher) and, in the process of classification, the assessed input (vector) is being classified on the basis of the principle of distance (Euclide's distance, most often) see (5). It is followed by the adjustment of its weights in the process of learning of this network. The number of sets into which the inputs will be classified is decided before the very calculation. For the purpose of an easy identification of the input data and its classification, based on common properties, the self-organizing networks will be used for the selection of an appropriate financial portfolio.

$$D = (x_1 - w_1)^2 + (x_2 - w_2)^2 + ... + (x_n - w_n)^2 \tag{5}$$

To perform the calculation itself, software for the construction of the self-organizing neural network model will be used [14]. The actual data, with regard to suppressing the influence of different input dimensions, will be adjusted according to need by standardization with the aim of integrating the weights of the assessed data [17]. The medium value of the standardized characters values then equals 0 and the variance is 1. During the process of learning of the neural network, suitable learning algorithms will be tested [21].

2.4 Input Data

As input data for verifying the access, data of capital funds with ESG, which take into account factors available on the Socially Responsible Investment organization website[6]. An advantage of this search-engine is the option of filtering funds based on the membership in a country where this fund is marketed. The employed data was filtered for Czech Republic (CR, 170 funds) and for the V4 (Visegrad Group, 380 funds) states, i.e., Czech Republic, Slovak Republic, Hungary and Poland. A partial view of this data can be seen in Tab. 1 (descriptive qualitative

[6] Socially Responsible Investment – https://yoursri.com/.

indicators). For further calculations we have used a risk-index – the Synthetic Risk Reward Indicator (the indicator acquires values of 0 to 7) and Performance for the selected period. If the trust fund was in another currency than the EURO, we performed a re-calculation into this currency according to the present exchange rate.

ID	Identifiers: ISIN	Investment Data: (SRRI)	Perfor-mance: 1 Year	Perfor-mance: 3 Year	Perfor-mance: 5 Year	Investment Data: Total Assets
x1	LU0963865323	4	-0.0205	-0.0034	0.0467	179 587
x2	LU0132414144	4	-0.0655	0.0302	0.0533	332 953 552
x3	LU0616241476	3	0.0004	0.0103	0.0343	18 918 800
x4	LU0119099819	3	-0.0117	0.0197	0.0275	66 715 200
…	…	…	…	…	…	…

Table 1 ESG Funds Data (descriptive indicators, selected funds, Czech Republic)

Further data describe the present title yield with the frequency of reading the data being one day. The data was acquired from the Morning Star[7] website. The input parameter for data acquisition was the fund identifier (ISIN). In both cases, we have used the means of automatic data processing in the form of the computer script. In Table 2, we can see the ESG funds data from Czech Republic (financial means in EURO) in the form of the title's yield change in the course of time.

ID	Date\ISIN	LU0963865323	LU0132414144	LU0616241476	LU0119099819	…
x1	27. Sep 2013	99.7457417	80.74965372	39.42816777	43.00642982	…
x2	28. Sep 2013	99.7457417	80.74965372	39.42816777	43.00642982	…
x3	29. Sep 2013	99.7457417	80.74965372	39.42816777	43.00642982	…
x4	30. Sep 2013	99.2292481	79.68452127	38.66715012	42.14741439	…
…	…	…	…	…	…	…

Table 2 ESG Funds Data – yield of the funds (selection, Czech Republic), 2013–2015

3 Results

Markowitz model

The Markowitz model was utilized in two sets of input data and this to SRI funds offered in Czech Republic and to SRI funds as part of the V4 group. The model assesses the expected profitability dependent on the risk of change. The investor chooses, from an acceptable portfolio set, a subset equal to the so-called "efficient" portfolio. This is a portfolio located on the so-called difference curve. On Fig. 1 this is a set of points through which we may pass a concave curve at its highest possible level. This curve begins in the positive values of the scales (see the color spectrum). Should the portfolio reach negative scales values, it is suitable to be sold. The model's output is a recommended degree of investment into the selected portfolio. As we can see from the graphic output (CR, V4), the available funds are suitable not only for conservative investors, but also for investors who prefer yield as well as the rising risk (top right part of the concave curve). For the calculation itself, we have used the Python tool for finance, namely the Investment Optimization Portfolio with Python [20].

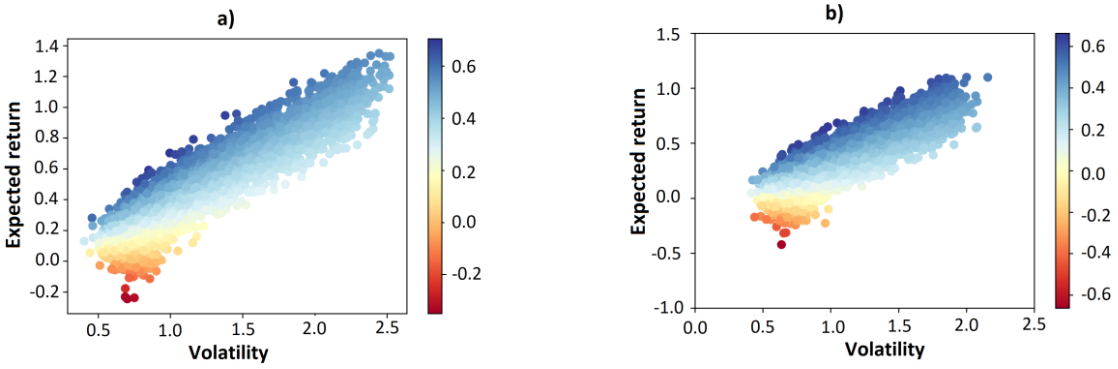


Figure 1 Markowitz model, a) Czech republic; b) GroupV4

[7] Morning Star – http://www.morningstar.co.uk/uk/.

Data Envelopment Analysis

The envelope method gives information about whether the given unit is efficient or not, vis á vis the efficiency boundary. As inputs, we have used data from Tab. 1, and this concretely the individual performance at input and Total Assets as the output parameter. The resulting situation for the collection of CR and V4 funds can be found in Fig.2. The method enables the contrast of the concrete fund with the other funds. The resulting suitable fund's portfolio would be chosen from the funds above the efficiency level.

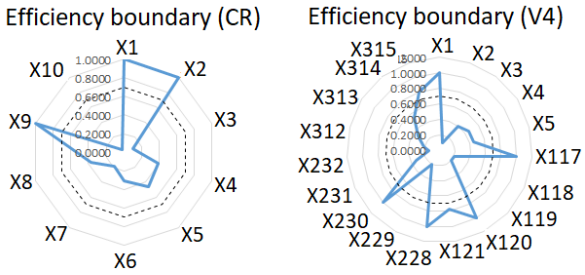


Figure 2 DEA – CR and V4 funds (selected funds)

Self-organizing neural network

Self-organizing neural networks were used for classifying input data (CR, V4), data in the structure according to Tab. 1, into sub-groups of funds that have similar characteristics. Emphasis was places on grouping according to the risk-factor, potentially also the sustainability level (ESG). Input data, therefore, was performance for 1, 3 and 5 years and furthermore, the ESG rating. The SRRI risk-factor coefficient was then taken as the control index. We managed to distribute the funds into four groups, see Fig. 3, while the groups could be ordered by ascending risk-factor as follows: r3, r2, r1, r4. The advantage of distributing the input data with the aid of the neural network model is the possibility of incorporating other parameters fairly easily – such parameters that have impact on the resulting distribution. For assessment, software developed as part of the research organization (self-organizing neural network) was utilized. for more details, see for example [14].

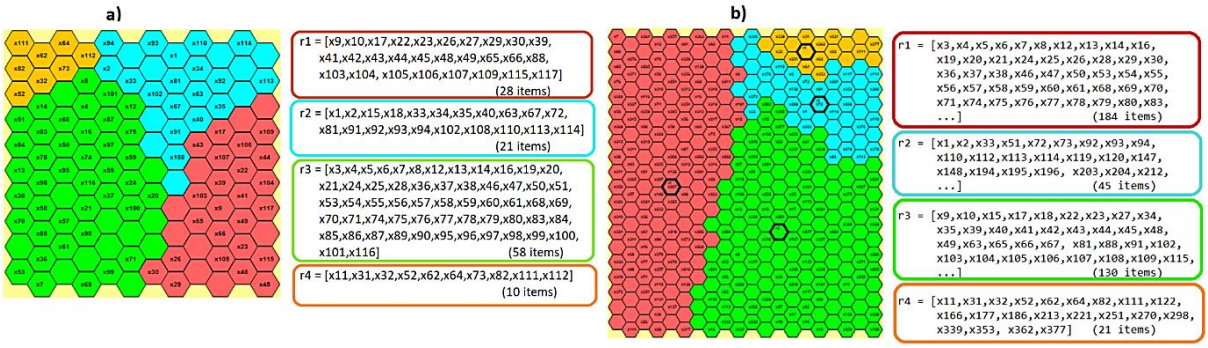


Figure 3 Self-organizing neuronal network, a) Czech republic; b) GroupV4

4 Conclusions

Selected methods were applied to input data with the aim of assessing funds, and this so that we could identify suitable funds for the investment portfolio. The criteira for funds selection were assessing the risk-factor and performance based on the availability of historical data in the selected period. Some investors are rather conservative; others are able to accept a higher risk with the expectation of higher returns. The Markowitz model is designated for rational investors who prefer higher returns and lower risk. [16]. An advantage is the ability to diversify the investment risk. As is visible in Fig.1, the character of funds in CR and V4, given suitable selection, ensures long-term returns. The DEA envelopes is suitable for assessing the efficiency of the production units, here the funds [7]. Thus the direction of funds that are efficient, on the border of efficiency, or unsuitable for further investments, see Fig. 2. Assessing the individual funds can be a method that is, in the comparison form, more lengthy. The most suitable method appears to be, for funds selection, self-organizing neural networks, where we are using the principle of proximity when assessing the input parameters and their subsequent classification into groups [14]. An advantage is the model's flexibility, the possibility of amplifying the classification criteria, and the simplicity of the model's integration into the implemented web-application.

Acknowledgements

The authors are thankful to the Grant Agency of Czech Republic (GAČR), project: "Modelling and simulation of sustainable investment decision-making", CEP ID GA17-23448S.

References

[1] Bengtsson, E. (2008) A History of Scandinavian Socially Responsible Investing, *Journal of Business Ethics*, 82(4), 969–983.
[2] Bonds and Climate Change. *The State of the Market Report 2016* (2016). [cit. 2019-4-21] Available at: https://www.climatebonds.net/files/files/CBI%20State%20of%20the%20Market%202016%20A4.pdf.
[3] Bugg-Levine, A., Emerson, J. (2011). Impact investing: transforming how we make money while making a difference. San Francisco: Jossey-Bass. p 306. ISBN 978-0-470-90721-4.
[4] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European-Journal of Operational Research*, 2, 429–444.
[5] Clark, G., Feiner, A., Viehs, M. (2015). *From The Stockholder to The Stakeholder*. University of Oxford. Oxford. p 63. [cit. 2019-4-21] Available at: https://arabesque.com/research/From_the_stockholder_to_the_stakeholder_web.pdf.
[6] Cowton, C. J., Sandberg, J. (2012). Socially responsible investment. *In Ruth Chadwick (ed.)*. Encyclopedia of Applied Ethics. Academic Press.
[7] Doyle, J.R., Green, R.H. (1995). Cross-evaluation in DEA: improving discrimination among DMUs. *INFOR 33*, 205–222.
[8] Duuren, E., Plantinga, A., Scholtens, B. (2016) ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *Journal of Business Ethic*. 138(3), 525–533.
[9] Edwin, J. (2007). *Modern portfolio theory and investment analysis*. John Wiley & Sons, p 728, ISBN 0470050829.
[10] Erste group. (2018). SRRI (Synthetischer Risiko- und Ertragsindikator). [cit. 2019-4-21]. Available at: http://fondsabc.erste-am.com/srri/.
[11] Forbes. (2017). *Socially-Responsible Investing: Earn Better Returns from Good Companies*. [cit. 2019-4-21] Available at: https://www.forbes.com/sites/moneyshow/2017/08/16/socially-responsible-investing-earn-better-returns-from-good-companies/#5542a29623d0.
[12] Hřebíček, J., Trenz, O., Chvátalová, Z., Soukopová , J. (2014) Optimization of Corporate Performance Using Data Envelopment Analysis with Maple. *In Aurelio Araujo et al. Engineering Optimization 2014*. London: CRC Press, 763–767. ISBN 978-1-138-02725-1.
[13] Jablonský, J. (2007). *Operations research: Quantitative models for economic decisions* (in Czech), Praha: Professional Publishing.
[14] Konečný, V., Trenz, O., Svobodová, E. (2010). Classification of companies with assistance of self-learning neural networks. *Agricultural Economics*. 56(2), 51–58.
[15] Markowitz, H. (1954) Portfolio Selection. *Journal of Finance*. [cit. 2019-4-21] Available at: http://www.gacetafinanciera.com/TEORIARIESGO/MPS.pdf.
[16] Markowitz, H. (1959) *Portfolio Selection: Diversification of Investments*. John Wiley & Sons, Inc., New York. [cit. 2019-4-21] Available at: http://cowles.econ.yale.edu/P/cm/m16/index.htm.
[17] Meloun M., Militky J. (1996) *Statistical processing of experimental data* (in Czech). Univerzita Pardubice, Czech Republic, ISBN 80-7194-075-5.
[18] Python. (2017). *Investment Portfolio Optimisation with Python*. [cit. 2019-4-21] Available at: http://www.pythonforfinance.net/2017/01/21/investment-portfolio-optimisation-with-python/.
[19] Russell, J., Norvig, P. (2010) *Artificial intelligence: a modern approach*. Upper Saddle River: Prentice Hall, ISBN 978-0-13-604259-4.
[20] Song, Y., Wang, H., Zhu, M. (2018). Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR. *Financial Innovation*. Springer. 4(2).
[21] Stencl, M., Stastny, J. (2010). Neural network learning algorithms comparison on numerical prediction of real data. *16th International Conference on Soft Computing MENDEL 2010*. Brno. 280–285.
[22] The Green New Deal. (2017). [cit. 2019-4-21] Available at: https://gpus.org/organizing-tools/the-green-new-deal/

Crossing numbers of join product of several graphs on 6 vertices with path using cyclic permutation

Emília Draženská [1]

Abstract.

The crossing number, $cr(G)$, of a simple graph G is the minimum number of edge crossings in a good drawing of G in the plane. In general, compute the crossing number for a given graph is a very difficult problem. The crossing numbers of a few families of graphs are known. One of them are join products of special graphs. Exact values of crossing numbers of the join products of graph G with discrete graph D_n, path P_n or cycle C_n are known for several graphs G. In the paper, we extend known results concerning crossing numbers for join of two graphs of order six with path P_n.

Keywords: graphs, drawings, crossing numbers, cyclic permutation, join product.

JEL classification: C02
AMS classification: 05C10; 05C38

1 Introduction

The problem of reducing the number of crossings was studied in many areas. The most interesting area is VLSI-layout. Introduction of the VLSI technology revolutionized circuit design and parallel computing. A lot of research concerning at efficient use of a new technologies has been done and further investigations are in progress. The crossing numbers has been also studied to improve the readability of hierarchical structures and automated graph drawings. The visualized graph should be easy to read and understand.

Let the *graph* G is a simple, undirected and connected with vertex set V and edge set E. A mapping that assings a point in the plane for each vertex and for each edge a continuous curve between its two endpoints is called a *drawing* D of the graph $G = (V, E)$. A *crossing* of two edges is the intersection of the interiors of the corresponding curves. The *crossing number*, $\mathrm{cr}(G)$, of a graph G is the minimum number of pairwise intersections of edges in any drawing of G in the plane. The drawing with a minimum number of crossings must be a *good drawing*, that means, each two edges have at most one point in common, which is either a commom end-vertex or a crossing.

Garey and Johnson proved [6] that computing the crossing number of a graph is an NP-complete problem. The exact values of crossings numbers are known for several special classes of graphs. One of them is a join products of two graphs. The join product $G_1 + G_2$ of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is obtained from the vertex-disjoint copies of G_1 and G_2 by adding all edges between $V(G_1)$ and $V(G_2)$. For $|V(G_1)| = m$ and $|V(G_2)| = n$, the edge set of $G_1 + G_2$ is the union of disjoint edge sets of the graphs G_1, G_2, and the complete bipartite graph $K_{m,n}$. Let D_n consists on n isolated vertices, let P_n be the paths with n vertices and C_n be the cycle with n vertices. In the proofs of the paper, we will often use the term "region" also in nonplanar drawings. In this case, crossings are considered to be vertices of the "map".

The exact values for crossing numbers of $G + P_n$ and $G + C_n$ for all graphs G of order at most four are given in [10], and the crossing numbers of the graphs $G + D_n$, $G + P_n$, and $G + C_n$ are also known for some graphs G of order five and six, see [2], [3], [4], [5], [9], [11], [12], [13], [15], [14], [16], and [17].

In this paper we extend these results by giving the exact values of the crossing numbers for join products for a special two graphs on six vertices with path P_n.

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, e-mail: emilia.drazenska@tuke.sk

Let D be a good drawing of the graph G. We denote the number of crossings in D by $\mathrm{cr}_D(G)$. Let G_i and G_j be edge-disjoint subgraphs of G. We denote the number of crossings between edges of G_i and edges of G_j by $\mathrm{cr}_D(G_i, G_j)$, and the number of crossings among edges of G_i in D by $\mathrm{cr}_D(G_i)$.

In the paper, it is used the Kleitman's result published in [8] on crossing numbers of complete bipartite graphs. More precisely, he proved that

$$\mathrm{cr}(K_{m,n}) = \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor, \quad \text{if} \quad m \leq 6. \tag{1}$$

2 The crossing number of $G + P_n$

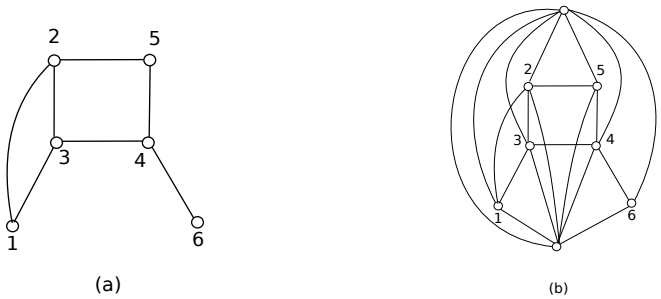


Figure 1: The graph G with the vertex notation and the graph $G + P_2$

We consider the join product of graph G (see Figure 1(a)) with path P_n on n vertices. The graph $G + P_n$ consists of one copy of the graph G and of n vertices $t_1, t_2, \ldots, t_n$, where any vertex t_i is adjacent to every vertex of G, $i = 1, 2, \ldots, n$, and also vertex t_i is adjacent to vertex t_{i+1}, $i = 1, 2, \ldots, n-1$. Let T^i, $1 \leq i \leq n$, denote the subgraph induced by the six edges incident with the vertex t_i and with vertex of G. Thus, $T^1 \cup T^2 \cup \cdots \cup T^n$ is isomorphic with the complete bipartite graph $K_{6,n}$ and

$$G + P_n = G \cup K_{6,n} \cup P_n = G \cup \left(\bigcup_{i=1}^{n} T^i \right) \cup P_n. \tag{2}$$

Let D be a good drawing of the graph $G + P_n$. The *rotation* $\mathrm{rot}_D(t_i)$ of a vertex t_i in the drawing D is the cyclic permutation that records the (cyclic) counter-clockwise order in which the edges leave t_i, see [7].

The rotation is a cyclic permutation. For $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$, every subgraph $T^i \cup T^j$ of the graph $G + P_n$ is isomorphic with the graph $K_{6,2}$. We will study the minimum number of crossings between the edges of T^i and the edges of T^j in a subgraph $T^i \cup T^j$ induced in D of $G + P_n$ depending on the rotations $\mathrm{rot}_D(t_i)$ and $\mathrm{rot}_D(t_j)$.

D. R. Woodall [18] published that in the subdrawing of $T^i \cup T^j$ induced by D is $\mathrm{cr}(T^i, T^j) \geq 6$ if $\mathrm{rot}_D(t_i) = \mathrm{rot}_D(t_j)$. And, if $Q(\mathrm{rot}_D(t_i), \mathrm{rot}_D(t_j))$ denotes the minimum number of interchanges of adjacent elements of $\mathrm{rot}_D(t_i)$ required to produce the inverse cyclic permutation of $\mathrm{rot}_D(t_j)$, then $Q(\mathrm{rot}_D(t_i), \mathrm{rot}_D(t_j)) \leq \mathrm{cr}_D(T^i, T^j)$.

We will separate the subgraphs T^i for $i, j \in \{1, 2, \ldots, n\}$ into three subsets. First, $R_D = \{T^i : \mathrm{cr}_D(G, T^i) = 0\}$. Second, $S_D = \{T^i : \mathrm{cr}_D(G, T^i) = 1\}$. And the last subset contains every T^i which crosses G at least twice in D.

Let F^i denote the subgraph $G \cup T^i$ for $T^i \in R_D$, where $i \in \{1, \ldots, n\}$. Thus, any F^i is represented by $\mathrm{rot}_D(t_i)$. All cyclic permutations of six elements can be generated using the algorithm published in [1].

We are interested only in such drawings of G if there is possibility to have $T^i \in R_D$. First, assume a drawing D of the graph $G + P_n$ in which the edges of G does not cross each other. In this case we have the vertex notation of G as it is shown in Figure 1 (a). So, in D there are only two possible configurations

of F^i (see Figure 2 and Table 1). Let $\mathcal{M}_\mathcal{D}$ denotes the set of configurations that exist in the drawing D belonging to $\mathcal{M}$.

Second, assume a good drawing D of the graph $G + P_n$ in which the edges of G cross each other exactly once. As there is $T^i \in R_D$, we have two possible drawings of G with a crossing (see Figure 3). And, the last possibility is a drawing in which there are at least two crossings on edges of G.

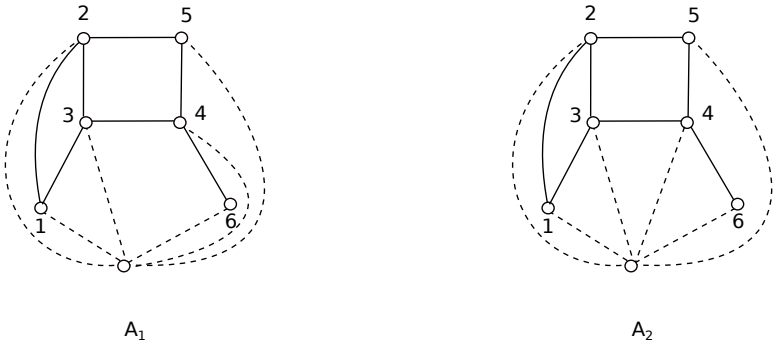


Figure 2: Drawings of all possible configurations of the graph F^i

$A_1 : (125643)$	$A_2 : (125463)$

Table 1: Configurations of graph $G \cup T^i$ with vertices denoted of G as in Figure 1(a)

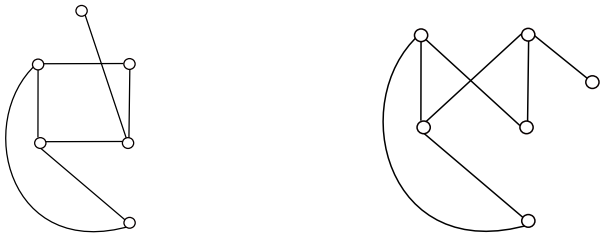


Figure 3: Two possible drawings of the graph G with one crossing

Let X, Y be the configurations from $\mathcal{M}_D$. We shortly denote by $\mathrm{cr}_D(X, Y)$ the number of crossings in D between T^i and T^j for different $T^i, T^j \in R_D$ such that F^i, F^j have configurations X, Y, respectively. Finally, let $\mathrm{cr}(X, Y) = \min\{\mathrm{cr}_D(X, Y)\}$ over all good drawings of the graph $G + P_n$.

Now, we compute the lower-bounds of number of crossing of configurations from $\mathcal{M}$. We have $\mathrm{cr}(A_i, A_i) \geq 6$ for $i = 1, 2$ and $\mathrm{cr}(A_1, A_2) \geq Q(\mathrm{rot}(t_i), \mathrm{rot}(t_j)) = 5$. It is summarized in Table 2.

$-$	A_1	A_2
A_1	6	5
A_2	5	6

Table 2: Lower bounds of numbers of crossings for two configurations from $\mathcal{M}$

Lemma 1. Let D be a good drawing of $G+P_n$, $n \geq 2$ with a vertex notation of the graph G as in Figure 1(a). If $T^n \in R_D$ such that F^n has configuration $A_i \in \mathcal{M}_\mathcal{D}$ for $i = 1, 2$, then $\mathrm{cr}_D(T^n, T^k) \geq 3$ for any $T^i \in S_D$.

Proof: Let in D the graph F^n has configuration A_1. If $T^k \in S_D$ with $\mathrm{cr}_D(T^n, T^k) = 2$, then the vertex t_k must be placed in a region with at least three vertices of G on its boundary, see Figure 2. Since $T^k \in S_D$, the vertex t_k cannot be placed in the region bounded by 4-cycle of the graph G. Moreover, if t_k is placed in another regions, then $\mathrm{cr}_D(T^n, T^k) \geq 3$. The same idea can be used for configuration A_2.□

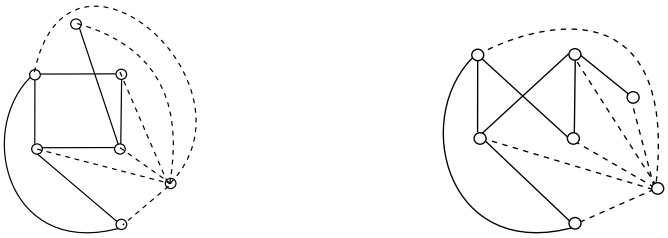


Figure 4: Two possible drawings of the graph F^n with one crossing

Theorem 1. $\mathrm{cr}(G+P_n) = 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1$ *for* $n \geq 2$.

Proof:

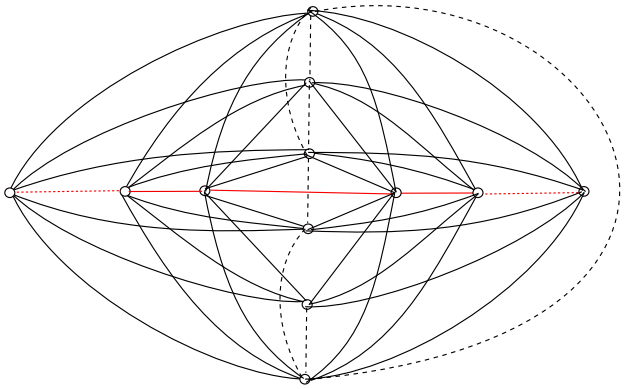


Figure 5: The graph $G + P_n$

There is a drawing of $G + P_n$ (see Figure 5) with $6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1$ crossings. Thus, we have $\mathrm{cr}(G+P_n) \leq 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1$. We prove the reverse inequality by induction on n. Using algorithm on the website `http://crossings.uos.de/`, we can prove that the result is true for $n = 2$. Suppose now that, for $n \geq 3$, there is a drawing D with

$$\mathrm{cr}_D(G+P_n) < 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1, \tag{3}$$

and let

$$\mathrm{cr}_D(G+P_m) \geq 6\left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor + 2\left\lfloor \frac{m}{2} \right\rfloor + 1 \qquad \text{for any integer } m < n. \tag{4}$$

As the graph $G + D_n$ is a subgraph of the graph $G + P_n$ and $\mathrm{cr}_D(G + D_n) = 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor$ (see [13]), then we assume that $\mathrm{cr}_D(G+P_n) = 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor$. Thus, no edge of the path P_n is crossed in D.

First, we prove that the considered drawing D must be antipodal-free, that is $\mathrm{cr}_D(T^i, T^j) \neq 0$ for all i, j. As a contradiction suppose that, without loss of generality, $\mathrm{cr}_D(T^{n-1}, T^n) = 0$. Since the graph $G \cup T^{n-1} \cup T^n$ contains $K_{4,3}$ as a subgraph, and $\mathrm{cr}(K_{4,3}) = 2$, we have $2 \leq \mathrm{cr}_D(G, T^{n-1} \cup T^n)$.

The fact that $\mathrm{cr}(K_{6,3}) = 6$ implies that any T^k, $k = 1, 2, \ldots, n-2$, crosses $T^{n-1} \cup T^n$ at least six times. So, for the number of crossings, in D, we have

$$\begin{aligned}
\mathrm{cr}_D(G+P_n) = \mathrm{cr}_D\left(G+P_{n-2}\right) + \mathrm{cr}_D(T^{n-1} \cup T^n) + \mathrm{cr}_D(K_{6,n-2}, T^{n-1} \cup T^n) + \\
+ \mathrm{cr}_D(G, T^{n-1} \cup T^n) \geq 6\left\lfloor \frac{n-2}{2} \right\rfloor \left\lfloor \frac{n-3}{2} \right\rfloor + 2\left\lfloor \frac{n-2}{2} \right\rfloor + 1 + 6(n-2) + 2 = \\
= 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 2.
\end{aligned}$$

It contradicts that D is not antipodal-free.

Moreover, our assumption on D together with $\mathrm{cr}(K_{6,n}) = 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor$ implies that

$$\mathrm{cr}_D(G) + \mathrm{cr}_D(G, K_{6,n}) \leq 2\left\lfloor \frac{n}{2} \right\rfloor.$$

Let us denote $r = |R_D|$ and $s = |S_D|$. Then,

$$\mathrm{cr}_D(G) + 0r + 1s + 2(n - r - s) \leq 2\Big\lfloor \frac{n}{2} \Big\rfloor.$$

Thus, $2r + s \geq 2n - 2\Big\lfloor \frac{n}{2} \Big\rfloor$. We will fixed one subgraph T^i.

Case 1: $\mathrm{cr}_D(G) = 0$.

We can choose the vertex notation of the graph as shown in Figure 1(a). At first, we prove, that $n \neq s$. If $n = s$, it means that for every $T^i \in S_D$ we have two possibilities.

(i) For every $i, j, i \neq j : \mathrm{cr}_D(T^i, T^j) \geq 3$.

Without loss of generality let $T^n \in S_D$ and let us fix $G \cup T^n$. Then using Lemma 1, we have

$$\begin{gathered}\mathrm{cr}(G + P_n) \geq \mathrm{cr}_D(K_{6,n-1}) + \mathrm{cr}_D(K_{6,n-1}, G \cup T^n) + \mathrm{cr}_D(G \cup T^n) \geq \\ \geq 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 4(n-1) + 1 > 6\Big\lfloor \frac{n}{2} \Big\rfloor \Big\lfloor \frac{n-1}{2} \Big\rfloor + 2\Big\lfloor \frac{n}{2} \Big\rfloor.\end{gathered}$$

(ii) There is $T^i, T^j \in S_D$, $i \neq j$: $1 \leq \mathrm{cr}_D(T^i, T^j) \leq 2$.

If there exists also such $T^k \in S_D$ that for every $T^l \in S_D$, $l \neq k$: $\mathrm{cr}_D(T^k, T^l) \geq 3$, we fixed $G \cup T^k$. And the same inequalities as in previous case (i) hold.

If there is not such $T^k \in S_D$, without loss of generality, let $1 \leq \mathrm{cr}_D(T^{n-1}, T^n) \leq 2$ and let us fix $G \cup T^{n-1} \cup T^n$. In this step we are interested in all possible configurations of the subgraph F^i for some $T^i \in S_D$. Using cyclic permutation it is possible to prove $\mathrm{cr}_D(G \cup T^{n-1} \cup T^n, T^i) \geq 7$ for every T^i. So, we have

$$\begin{gathered}\mathrm{cr}(G + P_n) \geq \mathrm{cr}_D(K_{6,n-2}) + \mathrm{cr}_D(K_{6,n-2}, G \cup T^n \cup T^{n-1}) + \mathrm{cr}_D(G \cup T^n \cup T^{n-1}) \geq \\ \geq 6\Big\lfloor \frac{n-2}{2} \Big\rfloor \Big\lfloor \frac{n-3}{2} \Big\rfloor + 7(n-2) + 3 > 6\Big\lfloor \frac{n}{2} \Big\rfloor \Big\lfloor \frac{n-1}{2} \Big\rfloor + 2\Big\lfloor \frac{n}{2} \Big\rfloor.\end{gathered}$$

Thus, $n \geq s + 1$. That implies $r \geq 1$. We assume that $T^n \in R_D$ with F^n having configuration A_1 or A_2. We will discuss two possibilities over congruence n modulo 2.

- Let n be even. By fixing the graph F^n and using Table 2 and Lemma 1, we have

$$\begin{gathered}\mathrm{cr}(G + P_n) \geq \mathrm{cr}_D(K_{6,n-1}) + \mathrm{cr}_D(K_{6,n-1}, G \cup T^n) + \mathrm{cr}_D(G \cup T^n) \geq \\ \geq 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 5(r-1) + 4s + 3(n - r - s) + 0 = \\ = 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 3n + (2r + s) - 5 > 6\Big\lfloor \frac{n}{2} \Big\rfloor \Big\lfloor \frac{n-1}{2} \Big\rfloor + 2\Big\lfloor \frac{n}{2} \Big\rfloor.\end{gathered}$$

- Let n be odd. By fixing the subgraph T^n,

$$\begin{gathered}\mathrm{cr}(G + P_n) \geq \mathrm{cr}_D(G + P_{n-1}) + \mathrm{cr}_D(G + P_{n-1}, T^n) \geq \\ \geq 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 2\Big\lfloor \frac{n-1}{2} \Big\rfloor + 5(r-1) + 3s + 1(n - r - s) + 0 = \\ = 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 2\Big\lfloor \frac{n-1}{2} \Big\rfloor + n + 2(2r + s) - 5 \geq \\ \geq 6\Big\lfloor \frac{n-1}{2} \Big\rfloor \Big\lfloor \frac{n-2}{2} \Big\rfloor + 2\Big\lfloor \frac{n-1}{2} \Big\rfloor + n + 2\Big(2n - 2\Big\lfloor \frac{n}{2} \Big\rfloor\Big) > \\ > 6\Big\lfloor \frac{n}{2} \Big\rfloor \Big\lfloor \frac{n-1}{2} \Big\rfloor + 2\Big\lfloor \frac{n}{2} \Big\rfloor.\end{gathered}$$

Case 2: $\mathrm{cr}_D(G) = 1$.

It means that $r \geq 1$. Without loss of generality, we assume that $T^n \in R_D$. We can choose the vertex notation of the graph as shown in Figure 1(a). In every drawing of the graph G with a crossing, it is possible to verify, $cr_D(G \cup T^n, T^i) \geq 4$ for every $i = 1, \ldots n-1$. So, by fixing the graph F^n we have

$$\mathrm{cr}(G + P_n) \geq \mathrm{cr}_D(K_{6,n-1}) + \mathrm{cr}_D(K_{6,n-1}, G \cup T^n) + \mathrm{cr}_D(G \cup T^n) \geq$$
$$\geq 6\left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 4(n-1) + 1 > 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor.$$

Case 3: $\mathrm{cr}_D(G) \geq 2$.

We use the same idea as in previous case for all possible drawings of the graph G with a possibility of an existence of a subgraph $T_i \in R_D$ in the considering D. In every such a drawing we obtain a contradiction. Thus, it completes the proof. □

The following theorem gives us the exact value of the crossing number of the graph $G + C_n$ for $n \geq 3$.

3 The crossing number of $H + P_n$

Let us consider the graph H (see Figure 6). As the graph $G + P_n$ is a subgraph of the graph $H + P_n$ and there is a drawing of the graph $H + P_n$ with $6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1$ crossings, the next result is clear.

Theorem 2. $\mathrm{cr}(H + P_n) = 6\left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2\left\lfloor \frac{n}{2} \right\rfloor + 1$ *for* $n \geq 2$.

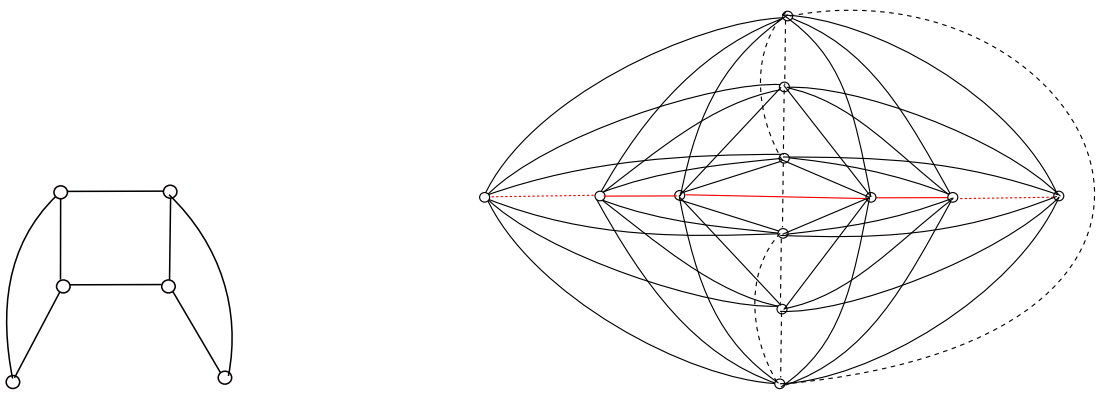


Figure 6: The graph H and the graph $H + P_n$

References

[1] Berežný, Š., Buša, J. Jr., Staš, M., *Software solution of the algorithm of the cyclic-order graph*, Acta Electrotechnica et Informatica, **18** (2018), No. 1, 3–10

[2] Berežný, Š., Staš, M., *On the crossing number of the join of five vertex graph G with the discrete graph* D_n, Acta Electrotechnica et Informatica, **17** (2017), No. 3, 27–32

[3] Berežný, Š., Staš, M., *Cyclic permutations and crossing numbers of join products of symmetric graph of order six*, Carpathian J. Math., **34** (2018), No. 2, 143–155

[4] Klešč, M., *The crossing number of join of the special graph on six vertices with path and cycle*, Discrete Math., **310** (2010), 1475–1481

[5] Klešč, M., *The join of graphs and crossing numbers*, Electron. Notes in Discrete Math., **28** (2007), 349–355

[6] Garey, M. R., Johnson, D. S., *Crossing number is NP-complete.* SIAM J. Algebraic. Discrete Methods, **4** (1983), 312–316

[7] Hernández-Vélez, C., Medina, C., Salazar, G., *The optimal drawing of $K_{5,n}$.* Electronic Journal of Combinatorics, **21(4)** (2014), 29

[8] Kleitman, D. J., *The crossing number of $K_{5,n}$.* J. Combinatorial Theory, **9** (1970), 315-323

[9] Klešč, M., Schrötter, Š., *The crossing numbers of join of paths and cycles with two graphs of order five*, Combinatorial Algorithms, Sprinder, LNCS, **7125** (2012), 160–167

[10] Klešč, M., Schrötter, Š., *The crossing numbers of join products of paths with graphs of order four*, Discuss. Math. Graph Theory, **31** (2011), 312–331

[11] Klešč, M., Valo, M., *Minimum crossings in join of graphs with paths and cycles*, Acta Electrotechnica et Informatica, **12** (2012), No. 3, 32–37

[12] Staš, M., *On the crossing number of the join of the discrete graph with one graph of order five*, J. Math. Model. and Geometry, **5** (2017), No. 2, 12-19

[13] Staš, M., *Cyclic permutations: Crossing numbers of the join products of graphs*, Proc. Aplimat 2018: 17^{th} Conference on Applied Mathematics, (2018), 979-987

[14] Staš, M., *Determining crossing number of one graph of order five using cyclic permutations*, Proc. Aplimat 2019: 18^{th} Conference on Applied Mathematics, (2019), 1126-1134

[15] Staš, M., *Determining crossing numbers of graphs of order six using cyclic permutations*, Bull. Aust. Math. Soc., **98** (2018), 353-362

[16] Staš, M., *Determining crossing number of join of the discrete graph with two symmetric graphs of order five*, Symmetry, **11** (2019), No. 2

[17] Staš, M., Petrillová, J., *On the join products of two special graphs on five vertices with the path and the cycle*, J. Math. Model. and Geometry, **6** (2018), 1-11

[18] Woodall, D. R., *Cyclic-order graphs and Zarankiewicz's crossing number conjecture*, J. Graph Theory, **17** (1993), 657-671

The influence of individual attributes in the supplier's choice of a capital spare part using AHP

Márcio Rodrigues[1], Eva Šírová[2]

Abstract. Decision making in the ongoing reality of digitalization of business processes in enterprises requires a quicker and more accurate analysis of a complex environment of information. In this scenario, current leaderships need the help of computational tools and structured multi-criteria decision making (MCDM) methodologies on a daily basis, such as the Analytic Hierarchy Process (AHP). This paper presents an application of this approach in the decision process between 3 suppliers of a capital spare part in a small industry. For this model, 3 decision makers directly involved in the purchasing process provided their judgments for the pairwise comparison between alternatives. They took into consideration 6 essential attributes related to the process, which were grouped into 2 hybrid criteria: Logistic-Economic and Quality-Sustainability. The goal of this paper is to bring a comparative overview of the final supplier ranking outputted from the model and its changes according to each individual attribute. These findings intend to raise a solid knowledge base of deeper aspects in key business processes, group decision making (GDM) and how essential MCDM models are to their understanding.

Keywords: Analytic Hierarchy Process (AHP), multi-criteria decision making (MCDM), spare parts, supply chain management, decision support systems.

JEL Classification: C440, M1
AMS Classification: 90B50

1 Introduction

In the professional literature, the most discussed topics related to spare parts management are classification and demand forecasting for stock control. However, the decision-making process dealing with the choice of the best supplier is also the focus of a number of research papers. In this context, the multi-criteria analysis approach, particularly the Analytic Hierachy Process (AHP), enables both qualitative and quantitative criteria to be considered in a classification/prioritization scheme, at the same time enabling the assignment of weights to the different parameters.

The aim of this paper is to create a model of the decision-making process of buying spare parts (SP) from different suppliers. This article presents a case study of the comparative application of AHP in a SME production company dealing with the development and production of system parts. For the decision between 3 suppliers for the purchase process of a capital spare part (SP) were asked three decision makers inherent in the process. For this model, 6 attributes (supplier lead times, availability, spare part unitary purchase price, part quality, aftersales customer service and sustainability) were grouped into 2 hybrid criteria (Logistics-Economic and Quality-Sustainability).

The contribution can be summarized as follows: The first section presents an overview of research on decision support models and the application of AHP technique in the business environment based on the results of the literature review. The second section describes the methodological steps followed within the creation of the model for the decision-making process of the supplier choice. Section 3 presents the results of the case study. It includes a brief company description, explanation of the chosen criteria and presents the hierarchy diagram of the AHP. The final section of the paper summarizes the results of the research and their possible limitations.

[1] Technical University of Liberec, Department of Business Administration and Management, Voroněžská 1329/13 460 01 Liberec, Czech Republic, marcio.rodrigues@tul.cz.
[2] Technical University of Liberec, Department of Business Administration and Management, Voroněžská 1329/13 460 01 Liberec, Czech Republic, eva.sirova@tul.cz.

2 Literature Review

AHP has been considered as a leading and one of the most popular multi-criteria decision-making techniques reported in the literature usually for classification or by the prioritization of alternatives. AHP uses a multi-level hierarchical structure of objectives, criteria, subcriteria, and alternatives. These comparisons are used to define the weight of each criterion and the relative performance measures of the alternatives for each criterion. This method also verifies the comparisons consistency and provides a mechanism to improve it in cases where the comparisons are not consistent [10].

In complex group decision-making, decision makers and decision attributes are the core of the relevant activities. A group decision-making method needs to assemble the preference information of the experts and/or decision makers (DMs) based on their knowledge and experience to produce consensus decision making results [5]. A structured and scientific decision-making process is essential to making a rational and applicable decision, especially in the dynamic business environment [4].

The other authors mention decision support models for supplier development have attracted increased attention in recent years. Research areas that belong to the field of supply chain management, such as identification of hidden critical suppliers in supply networks [8], approaches for green and sustainable supplier selection [1], supplier selection with fuzzy methods [6] have been discussed in recent years. Decision support models for supplier development can give valuable support in practice. For example by identifying suppliers suitable for development or by deriving optimal investment volumes for supplier development activities [2].

Regarding spare parts supply management, different techniques have been developed to enable the use of multicriteria models. There is a number of authors that deal with AHP and classification of spare parts [9]. The advantage of this technique is that it allows a weighted score and as such, it reduces a large number of criteria to a single variable with a consistent measure [7]. Maintenance and spare parts control strategy is the field of research of Hu [3].

3 Methods and objectives

The research methodology used to conduct this case study consists of the following steps:

- a brief literature review of recent applications of MCDM methods;
- identification of the motivation for this work and decision of using AHP and ANP as two relevant approaches;
- selection of the company and its current key decision process; the conception of the AHP model, criteria and attributes;
- interviews with involved decision makers; processing of the collected data and discussion of the results outputted from the model.

The main goals of this part of the article are the:

- identification of the production, maintenance and quality requirements;
- identification of the spare parts' criticality assessment criteria, based on the quality and logistic requirements;
- assignment of grades for each criteria, based on the maintenance requirements;
- assignment of grades for each criterion, based on the logistic requirements.

4 Case study

The company is dealing with high costs and unfeasible lead times in the purchasing process of spare parts, where most of its suppliers are not capable of delivering the needed goods on time and at a high cost, implicating on many production downtimes. Furthermore, many of the spare parts are critical to the production process, which are considered capital spare parts, mostly due to their high purchasing costs and supplier lead times. Especially for the capital parts, there are a few options of suppliers per parts, which makes this purchasing process even more challenging.

After an ABC analysis of the capital spare parts and ranked by the aforementioned factors which influence their criticality. Among others which will be further described in this research, the involved leadership have decided to improve the purchase process with the use of the proposed MCDM approach with its most critical capital part: the copper rods.

For this material, there are currently three suppliers from two different countries: one from Germany (hereafter mentioned as Supplier 1), and two from the Czech Republic, where one is from the Central Bohemian Region (Supplier 2) and the other from the North Bohemian Region (Supplier 3). Among all of them, Supplier 1 delivers the best quality of the parts but holds the most expensive price and it is the most distant from the company factory.

4.1 The decision makers, selected criteria and attributes

The maintenance department is the main responsible for the purchasing process of the copper rods. However, both production and quality departments also are part of it, where their managers meet weekly to review the future demand in order to assure production continuity. For this research and also agreed within the company that in order to the reliability of the proposed approach. The decision makers for this decision process were the maintenance, production and quality managers, respectively, as their departments have boundaries in the selected problematic. In order to detail and further discuss the results of this article in the next section, some information about the professional profiles of them is also presented.

4.2 The AHP hierarchy tree

Based on the previous literature and further investigation of recent applications of the AHP method for the supplier's choice decision, the authors found relevant criteria, which were used in previous researches.

During the conducted interviews with the aforementioned managers, these relevant factors that would influence the SP purchasing process were presented based on previous consolidated literature, and decided to choose 6 as attributes: delivery cycle time, supplier availability, price, part quality, aftersales customer service and sustainability.

Additionally, for this research the authors grouped them into two hybrid criteria. Mainly considering the feedback given by the managers regarding the relationship between the attributes: Logistic-Economic and Quality-Sustainability, as presented below in Figure 1.

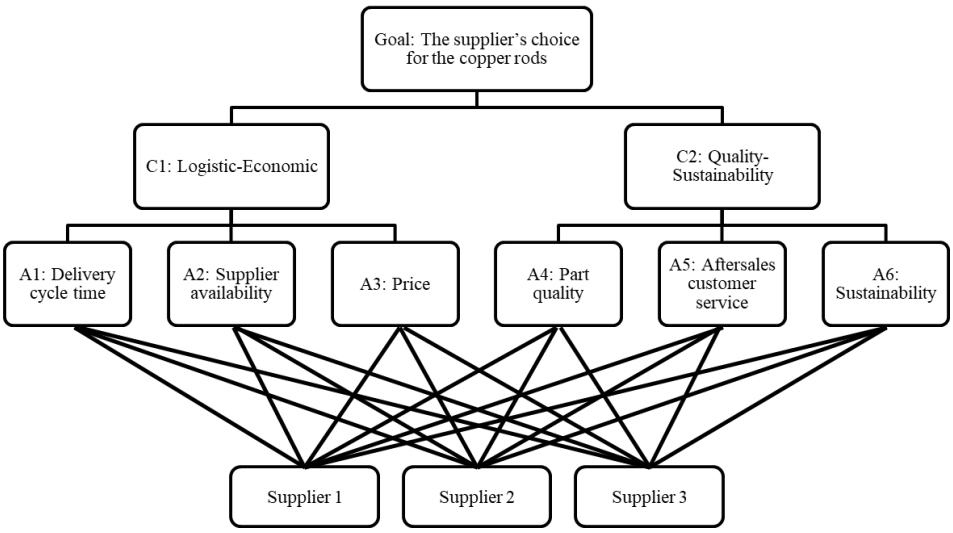


Figure 1 Hierarchical tree of the copper rods supplier's choice decision process

Taking into consideration that the AHP can also be exposed as a network approach, the previous model considers that both criteria have equal relative importance between each other. Therefore, the pairwise comparisons between supplier alternatives would take place at the attribute level only, as if they were criteria. In other words, it means that C1 is as important as C2, and consequently, A1, A2, A3, A4, A5, and A6 are equally important between each other.

5 Results and discussion

As previously mentioned in the methodological section of this paper, the authors conducted detailed interviews with the decision makers involved in this process and detailed information could be gathered. In order to structure the inputs from the managers, especially their judgments, an input formulary was created. Then all weights and grades were inputted into the software Super Decisions, by the Creative Decisions Foundation. An interesting tool

provided by the software is the information about inconsistency of judgments, one essential step in order to ensure the reliability of the outputs when using AHP method.

5.1 AHP Model Inputs

Considering the judgments of the three managers involved in the purchasing process, and using the geometric mean as previously mentioned at the beginning of the case study, the next tables extracted from the software Super Decisions shows the group evaluation between the attributes, as the relative importance between the criteria is 1.

In Tables 1 and 2, the judgments for the relative importance between attributes of each criteria are presented. Tables 3 presents the evaluation of the pairwise comparison between the three alternatives according to their performance in each attribute. In the next subsection, the main output obtained from these inputs using AHP is explored.

Logistic-Economic	Delivery Cycle Time	Price	Supplier Availability
Delivery Cycle Time	1	5	4
Price	1/5	1	1/3
Supplier Availability	1/4	3	1
Inconsistency	8.25%		

Table 1 Comparison and judgment for the hybrid criteria Logistic-Economic

Quality-Sustainability	Aftersales Customer Service	Part Quality	Sustainability
Aftersales Customer Service	1	1/3	2
Part Quality	3	1	3
Sustainability	1/2	1/3	1
Inconsistency	5.16%		

Table 2 Comparison and judgment for the hybrid criteria Quality-Sustainability

Criteria 1: Logistic-Economic				**Criteria 2: Quality-Sustainability**			
Delivery Cycle Time	Supplier 1	Supplier 2	Supplier 3	**Aftersales Customer Service**	Supplier 1	Supplier 2	Supplier 3
Supplier 1	1	1/5	1/3	Supplier 1	1	3	3
Supplier 2	5	1	3	Supplier 2	1/3	1	2
Supplier 3	3	1/3	1	Supplier 3	1/3	1/2	1
Inconsistency	3.70%			Inconsistency	5.16%		
Price	Supplier 1	Supplier 2	Supplier 3	**Part Quality**	Supplier 1	Supplier 2	Supplier 3
Supplier 1	1	1/6	1/8	Supplier 1	1	3	7
Supplier 2	6	1	1	Supplier 2	1/3	1	5
Supplier 3	8	1	1	Supplier 3	1/7	1/5	1
Inconsistency	0.89%			Inconsistency	6.24%		
Supplier Availability	Supplier 1	Supplier 2	Supplier 3	**Sustainability**	Supplier 1	Supplier 2	Supplier 3
Supplier 1	1	1/3	1/3	Supplier 1	1	3	5
Supplier 2	3	1	1	Supplier 2	1/3	1	3
Supplier 3	3	1	1	Supplier 3	1/5	1/3	1
Inconsistency	0.00%			Inconsistency	3.70%		

Table 3 Pairwise comparison judgement of suppliers according to each individual attribute

5.2 AHP Model Outputs

After computation of the group-consolidated evaluations of both criteria and attributes, the overall prioritization result between the three suppliers is presented in Table 4, denoting that the Supplier 2 is the recommended choice

to provide the copper rods to the company. This result had a big influence of some of the criteria: Delivery cycle time, Supplier Availability and Part Quality, calculated in the software Super Decisions.

Name	Normalized Weights	Rank
Supplier 1	0.371467	2
Supplier 2	0.419439	1
Supplier 3	0.209094	3

Table 4 Overall ranking among alternatives using the AHP

Analyzing the effects of the individual attributes and how they influence the final ranking, it is relevant to discuss this variation. Supplier 1 has the best performance in Part Quality, Aftersales Customer Service and Sustainability. Concerning Delivery Cycle Time, Supplier 2 has the best performance between all three alternatives, and also Supplier Availability. Supplier 3 also has the best Supplier Availability performance and the best price among all of three alternatives. Table 5 presents a comparative overview of the results, where the holder of the best performance in each attribute is highlighted.

Name	Delivery Cycle Time	Rank	Part Quality	Rank	Supplier Availabil-ity	Rank	Aftersales Customer Service	Rank	Price	Rank	Sus-taina-bility	Rank
Supplier 1	0.10473	3	0.64912	1	0.14286	3	0.59363	1	0.06724	3	0.63699	1
Supplier 2	0.63699	1	0.27895	2	0.42857	1	0.24931	2	0.44404	2	0.25828	2
Supplier 3	0.25829	2	0.07193	3	0.42857	1	0.15706	3	0.48872	1	0.10473	3

Table 5 Overall ranking among alternatives using the AHP

Furthermore, Supplier 2 was evaluated as the best provider of the copper rods according to the judgments of all three managers, regardless of the fact that it only had the best performance in two individual attributes, and both Supplier 1 and 3 had in three. So, theoretically, it would not be rated as the best choice in the whole model. This assumption would only make sense if all the weights in Tables 1 and 2 were equal to 1. However, the unequal importance between them pushed up Supplier 2 as the best option, mainly influenced by Delivery Cycle Time, Supplier Availability and Price. Figure 2 presents all ranking between alternatives in a comparative chart.

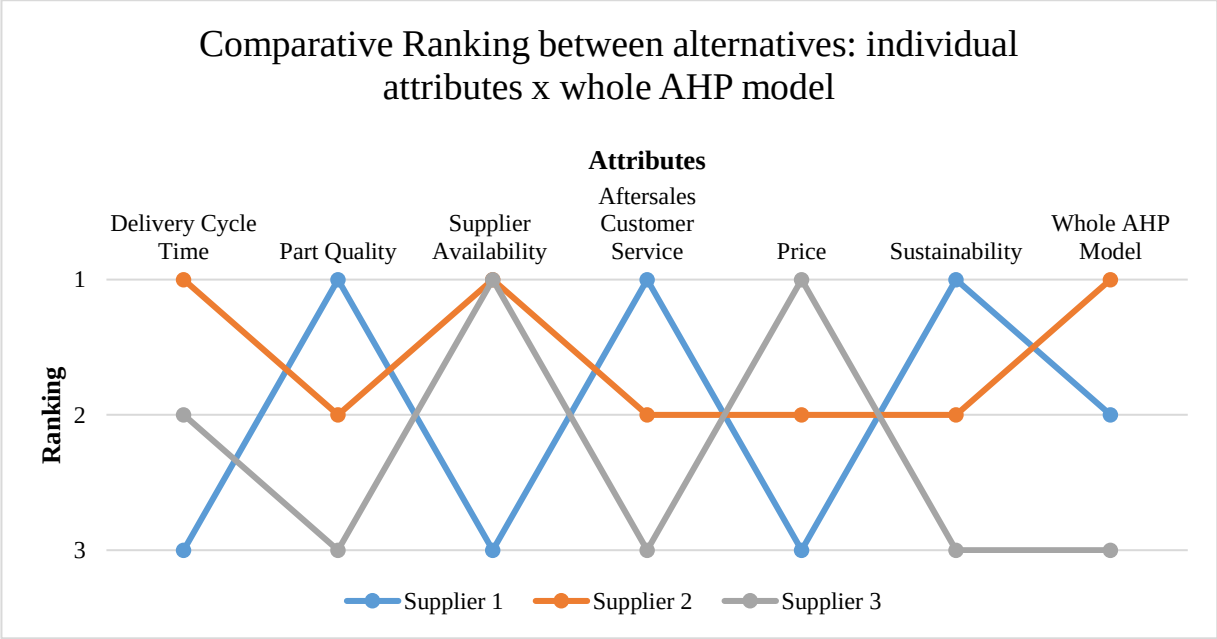


Figure 2 Comparative ranking between suppliers

6 Limitations and future research

One of the most difficulties faced by the authors during the interviews with the managers was convincing them to try a new mathematical approach to influence their decision process, where they initially refused to apply these two methodologies in their "sensible confidential data", which could affect their business strategy or other subjective aspects.

Another challenge was to deal with the consistency analysis of their evaluations, as they were using these methodologies for the first time, and, as the name of this research field suggests. These are frameworks intended to support decisions, not necessarily to determine them. Sometimes their judgments based on their former experience, regardless if it is mathematically inconsistent, can reflect in a deeper way the reality and particularity of the studied process better than a logical-structured set of steps in order to validate a relative preference comparison, here denoted as a decision.

7 Conclusion

This article presented a comparative application of the AHP methodology in the supplier's choice of a capital spare part. The initial section brought an overview of the current applications of those methods and a brief literature review. The case study presented the decision makers, their professional profiles and also the hierarchy and network frameworks for this application, including the criteria, attributes and their correlations. The next section discussed the results obtained from this research and also the evaluations obtained from the decision makers and how they influenced the outputs from both approaches.

Many senior leadership mainly trust their experiences and refuse to accept new methodologies. However, as previously mentioned in the introduction section of this article, the amount of data and raising complexity of the business processes in enterprises of all size demand the application of mathematical and computational tools to support decision-making, especially if there are many involved actors in the key process to optimize.

Acknowledgements

This article was supported by SGS 21301 "Project Management and Information Systems in Quality and Supply Chain Management" provided by Technical University of Liberec, Czech Republic.

References

[1] Alegoz, M., & Yapicioglu, H. (2019). Supplier selection and order allocation decisions under quantity discount and fast service options. *Sustainable Production and Consumption, 18*, 179-189.
[2] Glock, CH., Grosse, E. & Ries, J. (2017). Decision support models for supplier development: Systematic literature review and research agenda. *International Journal of Production Economics, 193*, 798-812.
[3] Hu, Q. (2018). OR in spare parts management: A review. *European Journal of Operational Research, 266 (2)*, 395-414.
[4] Koksalmis, E. & Kabak, O. (2019) Deriving decision makers' weights in group decision making: An overview of objective methods. *Information Fusion, 49*, 146-160.
[5] Liu, Y. (2019). An improved grey group decision-making approach. *Applied Soft Computing, 76*, 78-88.
[6] Rezaei, J., Fahim, P., & Tavasszy, L. (2014). Supplier selection in the airline retail industry using a funnel methodology: Conjunctive screening method and fuzzy AHP, *Expert Systems with Applications, 41(18)*, 8165-8179.
[7] Roda, I., Macchi, M., and Fumagalli, L. (2014). A review of multi-criteria classification of spare parts. *Journal of Manufacturing Technology Management, 25 (4)*, 528-549
[8] Shao, B. (2018). A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index. *Decision Support Systems, 114*, 37-48.
[9] Stoll, J., Koft, R., Schneider, J., & Lanza, G. (2015). Criticality analysis of spare parts management: a multi-criteria classification regarding a cross-plant central warehouse strategy. *Production Management, 9 (2*), 225- 235.
[10] Teixeira, C., Lopes, I. & Figueiredo, M. (2017). Multi-criteria Classification for Spare Parts Management: A Case Study. *Procedia Manufacturing, 11*, 1560-1567.

Local Action Groups in the Czech Republic and their assessment

Renata Klufová[1], Jana Klicnarová[2]

Abstract. The LEADER programme supports the endogenous and also neo-endogenous development of rural regions through a bottom-up and an integrated approach implemented by the Local Action Groups (LAGs). The evaluation of LAGs' effectiveness in achieving the programme's objectives has already been discussed by the scientific community. The aim of the LAGs evaluation in the Czech Republic is to identify factors that affect their success rate in obtaining subsidies.

Keywords: rural development, LEADER, LAGs, regression analysis

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The LEADER programme is a European Union initiative to support rural development projects initiated at the local level in order to revitalise rural areas and create jobs. It is implemented by around 2 600 Local Action Groups (LAGs), covering over 54% of the rural population in the EU and bringing together public, private and civil society stakeholders in a particular area. The main principle that LAGs should use when doing their activities is called the LEADER principle and is based on a "bottom-up" approach.

In the 2014 – 2020 programming period, the LEADER method has been extended under the broader term Community-Led Local Development (CLLD) to three additional EU Funds:

- the European Maritime and Fisheries Fund (EMFF);
- the European Regional Development Fund (ERDF);
- the European Social Fund (ESF)

Although LEADER is obligatory only under the EAFRD, a single action can now be supported under two or more of the four EU Funds at the same time through the concept of multi-funded CLLD. Where this is applied, it enables LAGs to comprehensively integrate local needs and solutions and helps to reinforce the links between rural, urban and fisheries areas.

In the Czech Republic, LEADER uses investment funds from the state budget on the basis of the Act on the State Budget of the Czech Republic for the relevant year. It is designed, on the one hand, by a local partnership of rural territories (LAGs) that develop and implement joint development strategies and intentions, and, on the other hand, local actors implementing specific projects. The LEADER CR program focuses primarily on new forms of improving the quality of life in rural areas, strengthening the economic environment and enhancing the natural and cultural heritage. The overall objective of the program is to improve the organizational capacity of entities operating in rural areas. At present, there are 179 LAGs in the Czech Republic, which cover almost 91% of the Czech territory and 58% of the population in 2017 (Figure 1).

1.1 Approaches to LAGs evaluation

Assessment of LAGs operation is not an easy task. The evaluation of LEADER approach cannot be implemented without framing the assessment into the larger societal discourse. According to Lopolito [2], LEADER indicates "how" to proceed rather than "what" needs to be done. The first analyses and evaluations of the activity of LAGs were undertaken in 2007, but these documents were largely descriptive [3].

The evaluations conducted by the Ministry of Agriculture, the owner of the LEADER methodology, which assigned points to LAGs based on a detailed questionnaire and a follow-up interview with the LAGs' representatives were of great importance. LAGs were classified, based on the total number of points they received, into four categories (A – D). The highest ranked LAGs were in the first category, while the lowest ranked LAGs were

[1] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, klufova@ef.jcu.cz
[2] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, janaklic@ef.jcu.cz

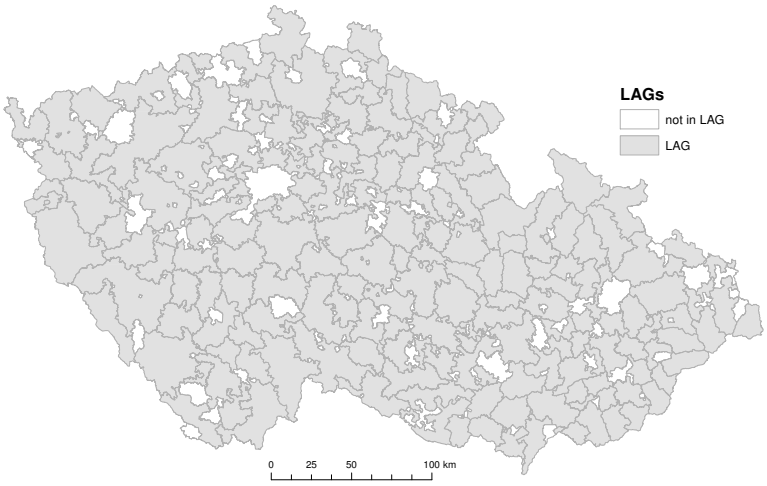


Figure 1 Area of the Czech republic covered by Local Action Groups

in the last category. The number of LAGs in the first two categories constantly increased while the number of LAGs in the C and D categories decreased – a positive trend. Additional attempts to assess the efficiency and performance of LAGs were to follow [5]. Activities of LAGs also are focused on strengthening relationships within local communities. One of the concepts that express these values is "community vitality" – a measure which takes into consideraton not only the economic health of a municipality but also at the quality of life in the municipality, the activities of non-profit organisations based there as well as the overall attractiveness of the municipality for its inhabitants [1]. In the Czech literature Vrabková & Šaradín [5] evaluate LAGs in terms of technical efficiency based on DEA model, Pechrová, A. & K. Boukalová [4] created a typology of the Czech LAGs according to their individual features and its organizational background.

Material and Methods

The aim of the LAGs evaluation in the Czech Republic is to identify factors that affect their success rate in obtaining subsidies. In our analysis we were looking for answers to the following questions:

1. Are the data provided by the Czech Statistical Office relevant to used subsidies resources?
2. Do the representatives of LAGs respond to demographic, economic and social conditions when they are preparing the projects?
3. Do the number of employees and the number of LAGs members affect the success rate in obtaining subsidies?

The analysis is based on data from publicly available resources: preliminarily subsidies allocation for individual LAGs[1] to date 14. 10. 2015 and other data for LAGs [2]. In the programming period 2014 – 2020 individual LAGs could prepare projects for obtaining grants from programs: Operational Programs - OP Employment and OP Environment, Integrated Regional Operational Program (IROP - Priority Axis 4) and Rural Development Program (RDP).

We were interested in the question which variables have an influence on allocated subsidies – especially, if the number of employees of the LAG office, the number of LAG members, the size of the LAG population, the number of completed dwellings, the number of economically active businesses, the rate of unemployment, the change of the unemployment rate (from 2008 to 2014), the proportion of seniors in LAG population, the change in this rate (from 2008 to 2014) and so on.

First, we studied the correlation matrix and because it is supposed that the non-relative measures are affected by the number of inhabitants, we calculated also the partial correlation matrix – where the variable the number of inhabitants in individually LAGs has been excluded.

In the second part of the analysis, we tried to explain the amount of allocated subsidies by characteristics for individual LAGs (the data had been aggregated from the municipal level to the LAG level). Since the data of the

[1] http://nsmascr.cz/aktualni-informace-pro-mas-v-op/
[2] https://www.czso.cz/csu/czso/data_pro_mistni_akcni_skupiny_mas

amount of allocated subsidies are for the year 2015, we used the data from the year 2014 and for some characteristics the change of rates between the years 2008 and 2014 – the LAGs had started their activities in the year 2008. The aim of LAG is to improve some characteristics in their region, hence we were interested in if the change in these characteristics affects the amount of allocated subsidies.

Therefore, we applied the regression analysis – we explained the amount of allocated total subsidies, resp. allocated subsidies in individual sub-programmes. The explanatory variables were data for individual LAGs as the number of inhabitants, the number of completed dwellings, the number of collective accommodation establishments, the change of unemployment rate, and the change of the proportion of seniors.

2 Czech LAGs performance evaluation - results and discussion

2.1 Correlation Analysis and Partial Correlation Analysis

From the correlation matrix, we can see that the allocated subsidies are statistically significantly correlated only with three relative variables – the change of relative seniors, with a rate of emigrants and the change of the unemployment rate. Since all other significant correlated variables are non-relative, it is supposed that the values are strongly affected by the number of inhabitants, so to study the relationship among them we apply the partial correlation matrix with excluded variable the number of inhabitants.

The results of partial correlations are quite surprising, it shows, for example, the significant negative correlation between the allocated subsidies and the number of registered economic subjects. It can be explained by the fact, that LAGs experiencing insufficient business activity on their territory (especially small and medium businesses), try to support activities of this type from different sources.

LEADER is the European Union's support program for rural development and support. Its specifics is that rural areas themselves distribute money through LAGs. They manage the co-operation of entrepreneurs, farmers, municipalities and non-profit organizations and share money with them. To be able to distribute money properly, the LAG must first and foremost have an idea of what it wants to support and why. What is the direction of their region? It must have prepared a so-called regional development strategy. These strategies usually include support for small and medium-sized enterprises in those LAGs where they appear to be inadequate.

Also, most of the relative indexes became significant so we can apply regression analysis in the following part. The signs of all significant correlations show that the subsidies are allocated into the parts with social problems (demographic ageing, the migration of younger people from municipalities and the associated reduction of human and social capital).

To check the relation among the number of employees, the number of members, and the amount of allocated subsidies, we use partial correlations with excluded variable number of inhabitants (since it is supposed that this variable has a significant influence on all these variables) and, surprisingly, there is no significant correlation. Therefore, we can deduce that there is no linear dependence between the amount of allocated subsidies and the number of employees, resp. the number of LAG members.

2.2 Regression analysis

If we evaluate the total pre-allocated amounts without distinguishing the source programs, we can see that statistically significant factors affecting the LAG's activity in order to obtain subsidies to support regional development were: LAG size (expressed in population), change in the age structure of the population (increase in seniors' share in ageing), changes in the unemployment rate, housing construction and the number of collective accommodation establishments in 2014 (see Table 2). The high value of the determination coefficient (see Table 1) confirms the assumption that the LAG is trying to use the available subsidy programs as much as possible and they adapt its development priorities to the priorities of the subsidy programs.

statistics	value
R	0.971449
R square	0.943713
adjusted R square	0.942096
Std. Error of the Estimate	11.245

Table 1 Total subsidies – Regression Statistics Table

The received positive impact of the size of LAGs on the amount of subsidies can be influenced by a number of factors - for example, in a larger population, the likelihood of higher levels of human and social capital, which are crucial to the existence and functioning of LAGs, is increased. The change in the age structure expressed by changes in the proportions of the three basic generations, which are mutually correlated variables, was reflected in the model by a change in the proportion of seniors. Given the fact that demographic ageing is accelerating and this will affect a wide range of aspects of human life and society, most municipalities and hence the LAGs are trying to prepare for these changes. These preparations can include a wide range of activities – from support for social services for seniors and families, leisure activities, seniors employment, barrier-free access to suitable forms of housing, etc. A range of activities or measures of this type can be funded from different programs. Numbers and changes in the number of collective accommodation establishments (hotels, pensions, etc.) were included in the analysis because tourism is one of the sectors currently supported in rural diversification. This is also confirmed by the correlation analysis, where changes in the number of collective accommodation establishments positively correlate with changes in the number of small and medium-sized businesses.

variable	coeff.	coeff. std.error	t	sig.
constant	-72.484	17.848	-4.0612	0.000074
O14	0.0025	0.00006	40.7519	0.000000
DB01-17	-0.0040	0.00088	-4.57028	0.000009
SENr08-14	76.5723	14.90927	5.13588	0.000001
NEZr08-14	-0.9416	0.12801	-7.35589	0.000000
HUZ14	0.0490	0.01581	3.09845	0.000074

Table 2 Total subsidies – Regression Coefficients Table

The indicator describing the total number of completed dwellings includes flats in both family and multi-dwelling buildings. It can, therefore, be assumed that the construction of family houses, which prevails in rural areas, is largely carried out by private persons and only a part is supported by the LAG in the framework of grant titles. The accuracy of the results of the analysis could be affected by the use of completed dwellings in 2008 – 2014, which is one of the topics for further refinement of the model. Similarly, the development of unemployment is expressed here by the development of the share of the unemployed in the total population, not in the economically active, which are not available at the municipal level. Given the record-low numbers of unemployed in recent years, it may be speculated that part of the required employment subsidies may thus concern pre-existing jobs that need to be maintained. However, this assumption needs to be verified by further detailed analysis.

Regression models for individual programs were also developed as part of the analysis. Allocated subsidies from the OP environment cannot be explained by the above-mentioned independent characteristics, which is very likely to be the objective of this program, where other variables would have to be included in the model. Models of other programs give satisfactory results.

Regression analysis – Integrated Regional Operational Programme

The quality of the model for subsidies from IROP is very similar to the model for total subsidies (see Table 1 and 3), which can also be influenced by the fact that the share of subsidies from this program in individual LAGs range from 41% to 84% (on average, LAGs gains 2/3 of subsidies just from this programme). LAGs can benefit from specific objectives[3] 4.1 - Strengthen community-led local development to improve the quality of life in rural areas and activate local potential, and 4.2 - Strengthening the capacity of community-led local development to improve management and administration skills of LAGs.

statistics	value
R	0.998417
R square	0.996837
adjusted R square	0.996746
Std. Error of the Estimate	1.645514

Table 3 IROP – Regression Statistics Table

[3] https://www.dotaceeu.cz/getmedia/a1500e9a-34af-4a8e-b6cb-8e0b119f26d4/Programming-document-of-IROP.pdf?ext=.pdf

variable	coeff.	coeff. std.error	t	sig.
constant	0.575	2.612	0.220	0.825952
O14	0.00155	0.000009	169.4355	0.000000
DB01-17	-0.0006	0.001	-4.716	0.000005
SENr08-14	5.435	2.182	2.491	0.013667
NEZr08-14	-0.111	0.019	-5.926	0.000000
HUZ14	0.008	0.002	3.426	0.000764

Table 4 IROP – Regression Coefficients Table

Within particular calls, LAGs can apply for subsidies for diverse activities[4] (social entrepreneurship, social infrastructure, cultural heritage, sustainable transport …). This is maybe another possible reason for the similarity of the models mentioned above. Therefore, it can also be stated that the above-mentioned comments on the model for subsidies in total can be applied here without any major problems (see Table 3).

Regression analysis – Rural Development Programme

The model developed for RDP subsidies shows a lower determination coefficient than the previous two models (Table 5). This is certainly influenced by the limitation of the programme for LAGs. The program supports the diversification of economic activities in the rural area in order to create new jobs and increase economic development[5].

statistics	value
R	0.877511
R square	0.770026
adjusted R square	0.763418
Std. Error of the Estimate	6.67780

Table 5 RDP – Regression Statistics Table

Community-led local development is supported, the LEADER method respectively, which contributes to better targeting of support for the local needs of the given rural area and the development of local actors' cooperation. The horizontal priority is the transfer of knowledge and innovation through educational activities and advice and cooperation in the field of agriculture and forestry. In the context of comprehensive rural development, the LAGs may, therefore, ask for support for specific activities.

variable	coeff.	coeff. std.error	t	sig.
constant	-28.847	10.560	-2.721	0.007159
O14	0.0007	0.00004	17.505	0.000000
DB01-17	-0.002	0.0005	-3.133	0.002028
SENr08-14	31.078	8.854	3.510	0.000571
NEZr08-14	-0.724	0.076	-9.524	0.000000
HUZ14	0.038	0.009	4.006	0.000091

Table 6 RDP – Regression Coefficients Table

However, the explanatory variables used in all models are reflected in this model in a similar way to the models discussed above. In contrast to expectations, the number of employees of the LAG's office and the size of their membership did not appear to be significant predictors in any of the models discussed. However, the correlation analysis and also partial correlation analysis performed shows a statistically significant positive dependence of both total subsidies and subsidies from IROP, OP Employment and RDP on both indicators. It may also be surprising that the model does not significantly determine the change of registered economic entities, especially small traders. This model is indirectly mediated by the number of collective accommodation establishments, as (as stated earlier), tourism is one of the sectors that are significantly supported in the diversification of activities in rural areas.

[4] see e.g. https://www.irop.mmr.cz/cs/Vyzvy/Detaily-temat/CLLD.
[5] https://www.szif.cz/cs/prv2014

Regression analysis – OP Employment

The motivation to obtain funds from the OP employment is very likely to be influenced by different factors. Using the same predictors as in the previous models, we get a significantly lower value of the determination coefficient (R^2 = 0.554). The variables collective accommodation establishments and unemployment were not significant. Therefore, other variables that could predict the indicated subsidy from this programme for each LAG were also tested: the crude rate of immigration, the share of seniors in 2014, the share of unemployed in 2014. However, the quality of the model did not increase (see Table 7).

variable	coeff.	coeff. std.error	t	sig.
constant	-13.8827	4.138415	-3.35459	0.000976
O14	0.0002	0.000018	9.92347	0.000000
DB01-17	-0.0009	0.000318	-2.81506	0.005439
SENr14	0.6448	0.187433	3.43996	0.000728
NEZr14	0.0815	0.01876	4.36479	0.000022
PRISTr14	0.1268	0.039863	3.17984	0.001745

Table 7 OP Employment – Regression Coefficients Table

The found predicting variables lead to the hypothesis that LAGs respond more to the actual situation in their micro-region when deciding to participate in this program.

3 Conclusions

Based on the analysis, it can be stated that the data provided by the CZSO are relevant and useful for statistical purposes of LAG performance, which was confirmed by the quality of the regression models created. Based on the results of regression models, it can also be assumed that LAG leaders respond to the challenges of the current demographic, economic and social situation of the micro-region. Confirming this hypothesis, however, would certainly deserve further detailed analysis at the level of individual LAGs. The calculations have also shown that the amount of subsidy received and allocated in particular LAGs correlated strongly with the LAG population size.

Negative partial correlation between the allocated subsidies and the number of registered economic subjects can be explained by the fact, that LAGs experiencing insufficient business activity on their territory (especially small and medium businesses), try to support activities of this type from different sources. There was no significant relationship, however, between the number of projects accepted in particular LAGs and the LAG membership size. This is an important finding which can be cited in debates concerning the importance of size for LAGs, in light of the fact that smaller LAGs are often regarded as less efficient. Similar finding obtained also other authors [5].

With regard to the scope of the contribution, it was not possible to discuss the issue in depth. However, there were further suggestions for a detailed analysis of the issue from the evaluation: refining the database for analysis, including in the assessment of the LAG the structure of their membership base; the number of challenges involved by the LAG (activity), LAG facilities, detailed analysis of projects by structure, etc. The discussed model was tested on preliminary allocations for 2015. In the next analysis, it will be appropriate to verify it in the refined allocations for 2018, which are unfortunately not yet available in the individual programs.

References

[1] Grigsby, W., J. (2001). Community vitality: Some conceptual considerations. *Rural Development Paper no. 6*, The Northeast Regional Center for Rural Development.

[2] Lopolito, A., Nardone, G. & R. Sisto (2011). Towards a comprehensive evaluation of Local Action Groups in LEADER programmes. *New Medit*, 1, 43–49.

[3] Pavlíková, G. (2007). Program LEADER ČR - nástroj pro rozvoj venkova. Praha: ČZU.

[4] Pechrová, A. & K. Boukalová (2015). Differences among Czech Local Action Groups in using selcted principles of LEADER. *Scientia Agriculturae Bohemica*, 46 (1), 41–48.

[5] Vrabková, I. & P. Šaradín (2017). The technical efficiency of Local Action Groups: A Czech republic case study. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 65 (3), 1065–1074.

Is it possible to use multi-criteria decision making methods for demographic ageing typology?

Jana Klicnarová[1], Renata Klufová [2], Marika Hrubešová[3]

Abstract. Demographic ageing is a global phenomenon. With trends continuing towards the contraction of working life, severe imbalances may occur in individual life cycles, in the structure of the workforce, and in socioeconomic provision for an ageing population. The paper focuses to demonstrate the usage of multi-criteria decision making methods for the typology of south bohemian rural municipalities according to its demographic ageing. PRIAM method and Aspiration-level interactive method is applied to chosen demographic and socio-economic criteria in order to describe the process of demographic ageing in South Bohemia.

Keywords: demographic ageing, South Bohemia, PRIAM method, Aspiration-level interactive method, GIS

JEL Classification: C44, J18
AMS Classification: 90C15

1 Introduction

An important demographic trend of the present is the ageing of the populations. The second half of this century will therefore belong to seniors. Life expectancy and at the same time lower births will result in a significant increase in the proportion of people over 65 in the population. In 2101, their share will rise according to the medium variant of the projection[1] up to 33% of the Czech population. In absolute terms, it will be 2.5 million seniors. This development is absolutely natural from a demographic point of view and is no different from that in Western or Northern Europe [2].

On the other hand, ageing in the demographic sense affects the entire population and, compared to an individual ageing, ageing populations can rejuvenate by increasing the proportion of young age groups. In contrast to individual ageing, which is derived from the chronological age of the individual, i.e. from the date of birth, we nowadays increasingly encounter the notion of **demographic ageing**. It is usually defined by the proportion of persons aged 65 and over in a given population, i.e. on the basis of the statistical population and the so-called retrospective age. In general, demographic ageing depends not only on prolonging human life and subsequently increasing the contingent of the senior population, but is often more affected by the volume of the child component, i.e. the birth rate. Therefore, we distinguish two types of demographic ageing:

- ageing of population at the summit (mortality dominated ageing) - increasing the share of older people due to longer life expectancy,
- ageing of population at the base (fertility dominated ageing) - reduction of younger age groups due to decreased fertility and birth rates.

Usually both types are running simultaneously. The fastest rate of growth in the share of the total population can then be observed for those aged 80 and over, the so-called "oldest old" [6] or as the "fourth age" [8].

Of course, the ageing population and the growing proportion of seniors will result in an increase in the economic burden on the pension system. Vlandas [7] shows that social democratic parties have more economically orthodox manifestos in European countries with more elderly people, and that the share of elderly is negatively correlated with inflation in both a sample of 21 advanced economies and a larger sample of 175 countries. While until 2008 the Czech Republic's pension account was in surplus every year, since 2009, given the economic crisis and the reduction in the premiums of some groups of the population, the total premium collected is lower than the sum of money needed to pay pensions. With the current unfavorable economic trends continuing, the deficit would

[1] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, janaklic@ef.jcu.cz
[2] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, klufova@ef.jcu.cz
[3] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, hrubesova@ef.jcu.cz
[1] Population projection of the Czech Republic up to 2100 - see https://www.czso.cz/csu/czso/projekce-obyvatelstva-ceske-republiky-do-roku-2100-n-fu4s64b8h4.

continue to rise, and by 2060, the amount needed to pay pensions would have already been about 60% higher than the total premium collected under current laws.

However, the economic consequences of an ageing population are also reflected in health care financing. It is understandable that older people need on average health care more frequently and to a greater extent than younger people. At the same time, the amount of health insurance premiums (which the state pays to the health insurance system) is lower than the premiums paid by people of working age. The ratio of health care costs to total premiums collected is growing, and this growth will continue, even if it is not as large as the growth in the burden on the pension system.

1.1 Demographic ageing in the South Bohemia

The development of the relative age structure of the South Bohemian Region is shown in the graph 1, which shows a continuous increase in the proportion of persons over 65 in the population. However, in the period 2008 – 2017, the share of the child component (0 – 14 years) was also gradually increased by 1.2%. The consequence of this development is a gradual decrease in the proportion of persons aged 15 – 64, i.e. persons in the working age.

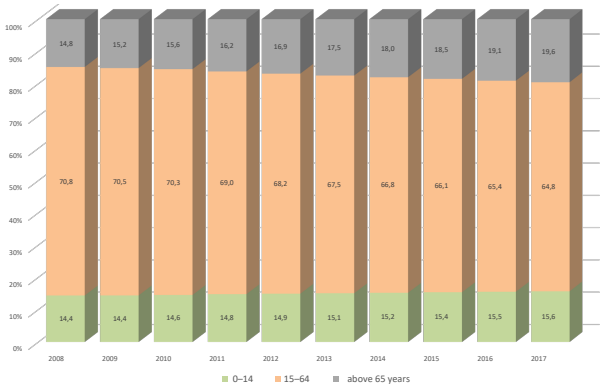


Figure 1 Development of the Region´s age structure 2008 – 2017

The share of seniors over 65 has increased since the beginning of the 21st century by 46.5% (from 85 404 in 2000 to 125 171 in 2017), with annual growth rising continuously. If the number of seniors increased by only 0.1% between 2000 and 2001, this was 2.7% between 2016 and 2017. The highest relative increase was recorded in 2010 – 2012, when the number of seniors in the South Bohemian population increased by 4.2% per year. These figures illustrate the increasing dynamics of absolute demographic aging.

If we look at the development of the child component of the population from the beginning of this century, the growth coefficients are considerably lower here. Between 2000 – 2007 the number of children in the South Bohemian Region decreased. They have been positive since 2008, but oscillate around 1%. Thus, the numbers of children in the population have been increasing since this year, but at a slower rate than those of seniors. The year-on-year relative changes in the numbers of children and seniors are shown in the graph 1.

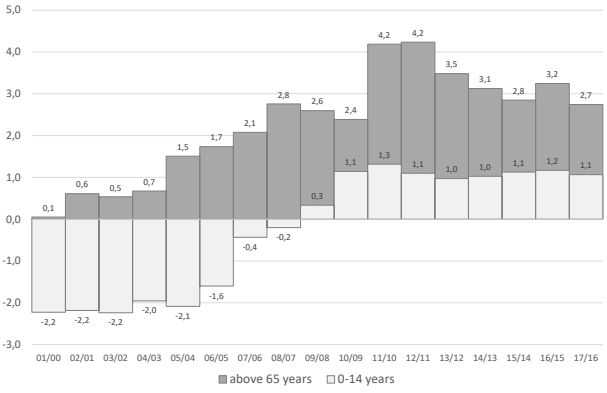


Figure 2 Year-on-year growth coefficients for children and seniors - 2000 – 2017 (%)

2 Material and Methods

The aim of the paper is to identify municipalities in the South Bohemian Region whose situation is unfavourable in terms of demographic ageing on one hand and on the other hand, to find municipalities whose position is relatively favourable. Thus, we will create a simple typology of the demographic ageing in the region. Regional demographic ageing policy should be then increasingly focused especially on the municipalities with unfavourable conditions. Thus, a set of basic characteristics describing this process was created. Rural municipalities with a population size of up to 3 000 inhabitants were included to an analysis (589 municipalities out of 624), as towns and cities have different characteristics, particularly in terms of availability and facilities of the necessary health and social care services. Demographic ageing is also assessed in connection with the so-called reproduction of the labor force. Boletice and Polná na Šumavě were excluded from the analysis (the first of them is a military area and the second one was established in 2016).

The following variables were chosen for the analysis: the change of the ageing index in 2008 – 2017 (IIS), the index of economic burden in 2017 (IHZ17), the mean age in 2017 (PV17), the change of the children's proportion in population 2008 – 2017 (IPD), the change of seniors proportion in population 2008 – 2017 (IPS), the so-called coefficient of labor exchange (describes the reproduction of the workforce) in 2017 - IVYM17, number of registered health and social care entities in 2017 at 1 000 inhabitants (SLUZBY), distance of the municipality from the nearest medical facility (NEAR_DOKT) and the distance of the municipality from the nearest social service for the elderly[2] (NEAR_SEN). The distances were calculated using ArcGIS 10.6.1 software. Detailed description of the indicators mentioned above can be found e.g. in [3].

The Aspiration-level interactive method, namely the conjunctive method and the PRIAM method, were used to analyze the municipalities of the South Bohemian Region. For a conjunctive method (for each maximizing, resp. minimizing) the criterion is set to a minimum value, respectively maximum value that each option must achieve. A variant is acceptable if it meets aspiration levels for all criteria at the same time. We then changed the aspiration levels, allowing the PRIAM method. Variables IIS, IHZ17, PV17, IPS, NEAR_SEN and NEAR_DOKT have been set as minimizing criteria, the other as maximizing. Particular variables can interact, so we grouped them according to their mutual relationships. The overall criterion for identyfing the most favourable and the most problematic municipalities can be expressed as follows:

$$(\text{IIS} \vee \text{IHZ17}) \wedge (\text{IPD} \vee \text{IPS}) \wedge \text{PV17} \wedge \text{IVYM17} \wedge [(\text{NEAR\_DOKT} \wedge \text{NEAR\_SEN}) \vee \text{SLUZBY}] \qquad (1)$$

Variables IIS and IHZ17 influence economic structure of the population, variables IPD and IPS inform about the size of dependent generations in the population. IPD refers to fertility dominated ageing, while IPS to mortality dominated ageing. IVYM17 describes the reproduction of the workforce (values above 1 indicate the so-called extended reproduction of the workforce, which leads to increasing numbers of people in the working age; values under 1 indicate the so-called reduced reproduction). The availability of health and social care is for an ageing population necessary.

3 Results and discussion

Basic demographic characteristics of the whole South Bohemia region were used as aspiration levels (Table 1), with an exception of distances to medical a social facilities, where medians were used. Demographic ageing is a complex problem connected with population demographic structure (age and gender structure) and processes taking place in society for a long time (economic, social, political). The current age structure reflects these processes.

Thus we identified 48 villages with the relatively favourable position according to demographic ageing in the region and 90 villages with unfavourable position (Figure 3). A comparison of the groups mentioned above provides Table 2, which consists of average values of the variables used in the analysis.

Table 2 shows differences between municipalities with favourable and unfavourable situation in demographic ageing. Municipalities with favourable position belong mostly to larger ones. We can distinguish two main subgroups among them:

1. suburban villages with the typical manifestations of the suburbanization process (moving younger, more educated people into these communities, higher birth rate, an intensive housing construction, commuting to nearby towns etc.). This is documented by many authors (e.g. [4], [5]);

[2] In line with the medium-term plan of social services of the South Bohemian Region for the period 2019 – 2020 the following types of services were selected in the category of services for the elderly: personal assistance, day care, emergency care, relief services, daily nursing homes, retirement homes, special regime homes and social services provided in institutional care facilities. For more see https://socialniportal.kraj-jihocesky.cz//?strednedoby-plan-rozvoje-socialnich-sluzby.

variable	aspiration level
mean age 2017	42.5
index of economic burden 2017	1.54
change of ageing index (%)	121.13
change of childern´s proportion 2008 – 2017 (%)	108.82
change of seniors proportion 2008 – 2017 (%)	131.85
coefficent of labor echxange	0.64
registered health and social care entities	0.75

Table 1 South Bohemia - aspiration levels

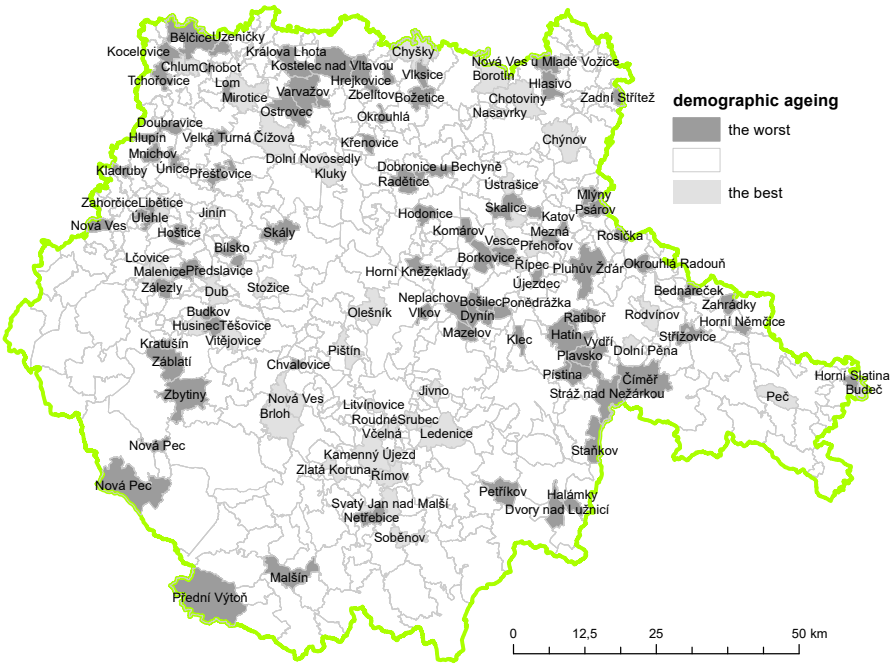


Figure 3 South Bohemian municipalities according to their position in the process of demographic ageing

variable	rural municipalities	the best	the worst
municipality size 2017 (inh.)	460	805	229
mean age 2017	42.5	39.8	45.6
index of economic burden 2017	1.55	1.51	1.58
change of ageing index (%)	123.7	96.1	152.3
change of childern´s proportion 2008 – 2017 (%)	116.5	116.4	97.4
change of seniors proportion 2008 – 2017 (%)	126.1	110.3	139.0
coefficient of labor exchange	0.71	0.90	0.42
registered health and social care entities	0.88	1.6	0.08
distance to the nearest health facility (km)	3.3	1.6	4.3
distance to the nearest seniors´ social care facility	5.6	3.7	6.3

Table 2 Demographic ageing in the South Bohemia region - groups comparison (average values)

2. villages with favourable location (e. g. in the transportation system) or with other factors that cause the site to develop.

This group shows on average a decrease in the age index in the reference period, indicating rejuvenation of the age structure. It corresponds to the lower mean age in this group. While the change in the proportion of children in the monitored period is comparable to the average for all rural municipalities, the average change in the share of seniors is significantly lower than in other municipalities. Suburban or a favourable transport accessibility is also confirmed by the distances to the nearest health and social facilities with services for seniors. Here it is worth mentioning the fact that the we used the Euclidean distances, not the distances corresponding to movement within the transport network. The relatively favourable demographic situation of the municipalities in this group is also documented by the average value of the index of economic burden and coefficient of labor exchange.

The group with unfavorable demographic aging characteristics is predominantly small-sized municipalities, either in peripheral locations or in less accessible locations of the region. The availability and size of these municipalities are also related to the availability of health and social care for the elderly. This is one of the problems of the South Bohemian Region, when health and social care becomes more expensive due to its size and the need of the workers to drive to relatively long distances. The values of the variables IPD a IPS in the table 2 confirm that both types of demographic aging are significantly manifesting here: the proportion of children in the population decreases and at the same time the share of seniors is growing significantly.

We can also look at some particular variables in a greater detail. From the economic point of view, the reproduction of workforce described by the coefficient of exchange IVYM17 is important for the future economic performance of the region. We have found 74 municipalities from the value of this coefficient greater than 1 which indicates an extended workforce reproduction. None of them belongs to urban municipalities (above 3 000 inhabitants), 9 of them belong to the group with identified favourable situation in demographic ageing. This group consists of small villages with the average population size of 316 inhabitants. Figure 4 shows their location which is important in the south bohemian settlement system.

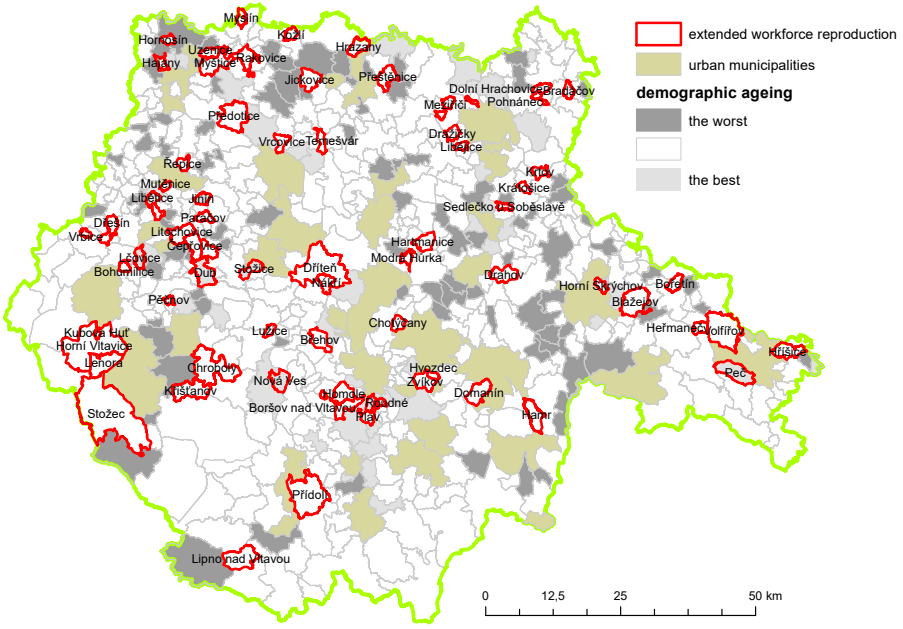


Figure 4 South Bohemian municipalities with extended reproduction of workforce

Some of these villages are located in the hinterland of towns or in a locations which are well accessible. But it couldn´t be a rule. Some of them are located in a peripheral or worse accessible locations. It could be connected with the process of amenity migration (see e. g. [1]) or its recreational use and intensive development of tourism industry (Stožec, Lipno nad Vltavou). The impact of migration on the development of the demographic structure cannot be ignored. However, because of its limited scope, this contribution was only devoted to natural increase. Verifying this hypothesis can be one of the topics for further detailed analysis. It should also be noted that it is appropriate to follow the evolution of the coefficient of exchange, not just its value in a given year.

4 Conclusions

In order to explore the processes involved in the population ageing intensively, a number of methods and techniques have been used. The new ones include an attempt to compare the number and proportion of major age categories. Some of them compare the pre-reproductive and post-reproductive categories of population and thus provide the knowledge about the development of the reproduction environment with respect to the exchange of generations. A set of indicators compares the productive population groups and provides information about the growing category of young age groups, as well as changes in the numbers of older age groups (retired persons). Mutual relations of these population ages introduce irreplaceable knowledge for the whole economic sphere. The quantitative relationship of "parents" and their "children" is important for the social welfare (security). Especially for the older age population groups such comparisons provide some idea of intergenerational support or care.

The contribution tried to use multi-criteria decision making methods for demographic ageing typology in South Bohemia region. The reason, why we used only these methods of multi-criteria decision-making is that we needed to compare the municipalities from the point of view of several criteria, however, we were not able to identify the importance of these criteria. Therefore, we could not use any methods where the weights play an important role. It was the main reason for the application of the chosen method. The Regional Authority is currently dealing with a project called Implementation of the Aging Policy at the Regional Level. The project includes demographic analysis. Identifying municipalities with unfavorable parameters of demographic aging can help to establish appropriate strategies for supporting seniors, developing health and social care, but also to support families with children. Using the Aspiration-level interactive method and PRIAM methods, we identified 48 municipalities with favourable characteristics of demographic ageing in suburban areas or localities with favourable location and 90 small municipalities with unfavourable demographic situation. Both identified groups show demographic characteristics that correspond in the literature discussed to the demographic development patterns of municipalities in relation to their population size, geographical location, economic and fiscal situation, etc.

The results indicate that the methods used can also be used to assess demographic aging and to evaluate other geographic processes that affect it. At present, the intensively discussed problem of demographic aging in the South Bohemian Region deserves further detailed analysis of the variables used, resp. inclusion of other variables (eg migration balance). Further attention should be paid to the choice of aspiration levels. The study of the issue can also be enriched by other indicators describing the issue of demographic aging in more detail (index of potential social support, dynamic population ageing metrics, dynamic reproduction ageing index). The process of demographic aging is also to be understood and described in the context of ongoing suburbanization processes, counter-urbanization, resp. amenity migration or rural gentrification.

References

[1] Barotoš, M. & others (2011). *Amenitní migrace do venkovských oblastí Český republiky*. Kostelec nad Černými Lesy: Nakladatelství Lesnická práce, s.r.o.

[2] European Commission. (2018). *The 2018 Ageing Report. Economic Budgetary Projections for the 28 EU Member States (2016–2070)*. Brussels: Directorate-General for Economic and Financial Affairs. Institutional Paper 079.

[3] Klufová, R. & Z. Poláková (2010). *Demografické metody a analýzy: demografie české a slovenské populace*. Praha: Wolters Kluwer.

[4] Muller, J. (2010). Demografické změny a jejich územní souvislosti, posilování či oslabování měst, suburbánní rozvoj a jeho diferencovaná dynamika. *Urbanismus a územní rozvoj*, XIII (5), 31–39.

[5] Ouředníček, M.& Špačková, P. & J., Novák (2013). *Sub urbs: krajina, sídla a lidé*. Praha: Academia.

[6] Rychtaříková, J. (2010). Demografické stárnutí: společný problém, stejné strategie? *Zpravodaj České demografické společnosti*, 17 (51), 1–2.

[7] Vlandas, T. (2017). "Grey power and the economy: Ageing and inflation across advanced economies". *Comparative Political Studies*, 51 (4): 514–552.

[8] Weeks, J., R. (2011). *Population: An Introduction to Concepts and Issues*. Belmont: Cengage Learning.

Comparison of similarity measures for generalized trapezoidal fuzzy numbers

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. In the last ten years, several new similarity measures for generalized trapezoidal fuzzy numbers were proposed in the literature. These similarity measures differ in the combination of properties of fuzzy numbers that are taken into consideration (e.g. center of gravity, height, area, perimeter, distance of significant values, etc. of the compared fuzzy numbers). The performance of these similarity measures was, so far, investigated only on several specifically chosen examples. A thorough comparison of the performance of these different similarity measures is still not available in the literature. This paper investigates the relationships between the new similarity measures using a numerical experiment. The effect of the different heights of the compared generalized fuzzy numbers is also considered. As such, the paper provides first insights into the shared features of the similarity measures.

Keywords: Generalized fuzzy numbers, similarity, numerical experiment, analysis

JEL Classification: D81, C44
AMS Classification: 90B50, 91B06

1 Introduction

The distance or similarity of fuzzy numbers is an important notion in fuzzy evaluation and fuzzy control. It is crucial for the ability to order fuzzy numbers and to decide the most appropriate linguistic approximation for the fuzzy number (among other uses). New distance and similarity measures of fuzzy numbers are still being proposed. Although there are several similarity measures proposed in the literature in the last ten years (see section 3), the authors of these methods provide only brief comparison of these measures and state their motivation for the newly proposed ones by examples. These comparisons are solely based on the comparison of the results of each considered similarity measures obtained for several pairs of generalized trapezoidal fuzzy numbers (the number of pairs ranges between 15 and 32 in [9, 10, 2, 5, 3]). Despite the fact that some more extensive investigations of the properties of similarity measures were provided e.g. in [7, 6, 8], a thorough comparison of the performance of similarity measures is, to our knowledge, not available. The goal of this paper is to provide more insights into the properties of these similarity measures and to identify their relationships using a numerical experiment. All this in the hope of understanding the added value of the newly proposed similarity measures over the already existent ones. For this reason, we have chosen a series of similarity measures, each inspired to some extend by previously introduced ones and by the inability of the previous once to distinguish between specific generalized fuzzy numbers.

2 Preliminaries

Let U be a nonempty set (*the universe of discourse*). A *fuzzy set* A on the universe U is defined by the mapping $A : U \to [0, 1]$. A family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$. For each $x \in U$ the value $A(x)$ is called the *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. Let A be fuzzy set on the same universe U. The set $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes the *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-*cut* of A for any $\alpha \in [0, 1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. $\mathrm{Hgt}(A) = \sup\{A(x) | x \in U\}$ denotes a height of fuzzy set.

A *fuzzy number* is a fuzzy set A defined on the set of real numbers which satisfies the following conditions: $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*); A_α are closed intervals for all $\alpha \in (0, 1]$; and $\mathrm{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on $[a, b] \subset \mathbb{R}$, if $\mathrm{Supp}(A)$ is a subset of the interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are

[1] Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.
[2] Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$.

The fuzzy number A is called *trapezoidal* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$ and $a_1 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A \sim (a_1, a_2, a_3, a_4)$. Fuzzy set A_G on U is called *generalized trapezoidal fuzzy number* if there exists a trapezoidal fuzzy number A and $w_A \in [0, 1]$ for which $A_G(x) = w_A \cdot A(x)$, $x \in U$. Generalized trapezoidal fuzzy numbers will be denoted by $A \sim (a_1, a_2, a_3, a_4; \mathrm{hgt}(A))$. For more information about fuzzy sets see e.g. [4].

3 Studied similarity measures for generalized trapezoidal fuzzy numbers

In this paper we will study the relationship between five similarity measures for generalized trapezoidal fuzzy numbers that were proposed in the last 10 years. These similarity measures are based on the similarity measure proposed in [1]. Each of them takes into consideration different combination of properties of generalized trapezoidal fuzzy numbers such as center of gravity, height, area or perimeter of the generalized fuzzy number the and distance of significant values.

Let $A \sim (a_1, a_2, a_3, a_4; \mathrm{hgt}(A))$ and $B \sim (b_1, b_2, b_3, b_4; \mathrm{hgt}(B))$ be generalized trapezoidal fuzzy numbers on $[0, 1]$, then the studied similarity measures are:

- *similarity measure* s_1 proposed by Wei and Chen in 2009 [9]:

$$s_1(A, B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\} + \min\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}{\max\{Pe(A), Pe(B)\} + \max\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}, \quad (1)$$

  where $Pe(A) = \sqrt{(a_1 - a_2)^2 + \mathrm{hgt}(A)} + \sqrt{(a_3 - a_4)^2 + \mathrm{hgt}(A)} + (a_3 - a_2) + (a_4 - a_1)$ is the perimeter of A, $Pe(B)$ is defined analogously.
- *similarity measure* s_2 proposed by Wen, Fan, Duanmu and Yong in 2011 [10]:

$$s_2(A, B) = (1 - |X_A - X_B|) \cdot (1 - |\mathrm{hgt}(A) - \mathrm{hgt}(B)|) \cdot \frac{\min\{Pe(A), Pe(B)\} + \min\{Ar(A), Ar(B)\}}{\max\{Pe(A), Pe(B)\} + \max\{Ar(A), Ar(B)\}}, \quad (2)$$

  where $Ar(A) = \frac{1}{2} \cdot \mathrm{hgt}(A) \cdot (a_3 - a_2 + a_4 - a_1)$ is area of A, $Ar(B)$ is computed analogously, $Pe(A)$, $Pe(B)$ are computed identically as in s_1 and X_A are the coordinates of the center of gravity of fuzzy number A calculated using the following formulas:

$$X_A = \frac{Y_A \cdot (a_3 + a_2) + (a_4 + a_1) \cdot (\mathrm{hgt}(A) - Y_A)}{2 \cdot \mathrm{hgt}(A)}, \text{where } Y_A = \begin{cases} \frac{\mathrm{hgt}(A) \cdot \left(\frac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6}, & \text{if } a_4 \neq a_1 \\ \frac{\mathrm{hgt}(A)}{2}, & \text{if } a_4 = a_1 \end{cases} \quad (3)$$

  and X_B are coordinates of the center of gravity of B defined analogously.
- *similarity measure* s_3 proposed by Hejazi, Doostparast and Hosseini in 2011 [2]:

$$s_3(A, B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\}}{\max\{Pe(A), Pe(B)\}} \cdot \frac{\min\{Ar(A), Ar(B)\} + \min\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}{\max\{Ar(A), Ar(B)\} + \max\{\mathrm{hgt}(A), \mathrm{hgt}(B)\}}, \quad (4)$$

  where $Pe(A)$, $Pe(B)$, $Ar(A)$ and $Ar(B)$ are computed identically as in the s_2.
- *similarity measure* s_4 proposed by Patra and Mondal in 2015 [5]:

$$s_4(A, B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right) \cdot \left(1 - \frac{1}{2} \cdot (|Ar(A) - Ar(B)| + |\mathrm{hgt}(A) - \mathrm{hgt}(B)|)\right), \quad (5)$$

  where $Ar(A)$ and $Ar(B)$ are computed identically as in the s_2.
- *similarity measure* s_5 proposed by Khorshidi and Nikfalazar in 2017 [3]:

$$s_5(A, B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4} \cdot d(A, B)\right) \cdot \left(1 - \frac{|Ar(A) - Ar(B)| + |\mathrm{hgt}(A) - \mathrm{hgt}(B)| + \frac{|Pe(A) - Pe(B)|}{\max\{Pe(A), Pe(B)\}}}{3}\right), \quad (6)$$

  where $d(A, B) = \frac{\sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2}}{\sqrt{1.25}}$, $\frac{|Pe(A) - Pe(B)|}{\max\{Pe(A), Pe(B)\}} = 0$ when $\max\{Pe(A), Pe(B)\} = 0$, $Pe(A)$, $Pe(B)$, $Ar(A)$ and $Ar(B)$ are computed identically as in s_2 and $[X_A, Y_A]$ and $[X_B, Y_B]$ representing the center of mass of fuzzy numbers A and B are computed using (3).

4 Numerical experiment

The investigation of the relationship of similarity measures presented in the previous section is based on the ideas used in the papers where these measures were proposed. In all five papers the authors selected several pairs of generalized trapezoidal fuzzy numbers and calculated the similarity of these pairs using several similarity measures. These results were then investigated and discussed. This was done to show a specific discrimination capability of the newly introduced similarity measure. Unfortunately a general comparison of the performance of the newly introduced measure with the already available ones was not performed in any of the papers. This prevents a reasonable comparison of the overall performance of these similarity measures. This paper aims on providing such a comparison.

To provide a thorough investigation, we conduct a numerical experiment that may shed more light on the relationship between the similarity measures. In this we will restrict ourselves only on trapezoidal fuzzy numbers (the examination of the relationships between the selected similarity measures for generalized trapezoidal fuzzy numbers will be a subject of future research).

One million pairs of trapezoidal fuzzy numbers (A_i, B_i), $i = 1, \ldots, 1000000$ was generated. After this each similarity measure $s_1, \ldots, s_5$ was applied on each pair. The values of the computed similarities were subsequently compared pairwise (and plotted using scatter plots for each pair of similarity measures). The correlations of each pair of similarity values were calculated. The results are depicted in Figure 1. The main diagonal represents plots of histograms of the actual values of each similarity obtained for the one million randomly generated pair of trapezoidal fuzzy numbers.

The analysis summarized in Figure 1 allows us to compare the results of the use of each similarity with all the other similarities in terms of their values. Pearson correlation coefficients are also calculated. From Figure 1 we can conclude several observation about the similarity measures:

- All investigated similarity measures are highly correlated. The lowest obtained value of correlation coefficient is 0.91 (obtained for measures s_1 and s_5). On the other hand the highest value of correlation coefficient was obtained for measures s_1 and s_4 and is equal to 0.99. The high correlation is not suprising given the fact that all of the similarity measures use the significant values in their formulas (with exception of s_2). Also the evolution from s_1 to s_5 proceeded in subsequent modifications of the previous versions of the similarity measures. These modifications were usually motivated by the inability of the previous similarity measure to distinguish between specific types of fuzzy numbers.
- Similarity measures s_1 and s_4 exhibit similar behaviour (see the extremely high correlation). This behaviour is unexpected because s_1 takes into account perimeters and s_4 areas of trapezoidal fuzzy numbers (remaining properties considered by s_1 and s_4 are the same: the height of fuzzy numbers and the distance of significant points). This can be interpreted in terms of the perimeter and the area of a trapezoidal fuzzy number carrying almost the same information about the fuzzy number. Note that this may no-longer hold for generalized trapezoidal fuzzy numbers.
- The distribution of the values of similarity measure s_5 is skewed to the right and its mean value is close to one, in other words the values of similarity of two trapezoidal fuzzy numbers might be the most inflated using s_5 (out of the five analysed similarity measures). This can be partially explained by the fact, that s_5 emphasizes the height of the generalized fuzzy number - the measure considers not only the height of generalized fuzzy number, but also their center of mass.
- s_3 and s_5 are also highly correlated (0.97), but the values of s_3 are more evenly spread over $[0, 1]$ interval.
- Histograms of all measures are skewed to the right with the exception of s_3, where the effect is not as strong.

Overall, each of the analysed similarity measures seems to be bringing something new to the table, while preserving the original idea of s_1.

5 Conclusion

The paper summarizes several recently proposed similarity measures of generalized trapezoidal fuzzy numbers and proposes a numerical experiment that allows us to understand the relationship between these measures better. The numerical experiment is based on the ideas used in the papers where the measures were proposed, but instead of a deliberate selection of specific pairs of generalized trapezoidal fuzzy numbers we performed a more straightforward approach based on a random generation followed by a visualization of the correlations among the similarity measures.

Despite the fact that all five similarity measures are highly correlated, we have succesffully identified several properties and relationships between these measures. It is apparent, that the modifications proposed in time

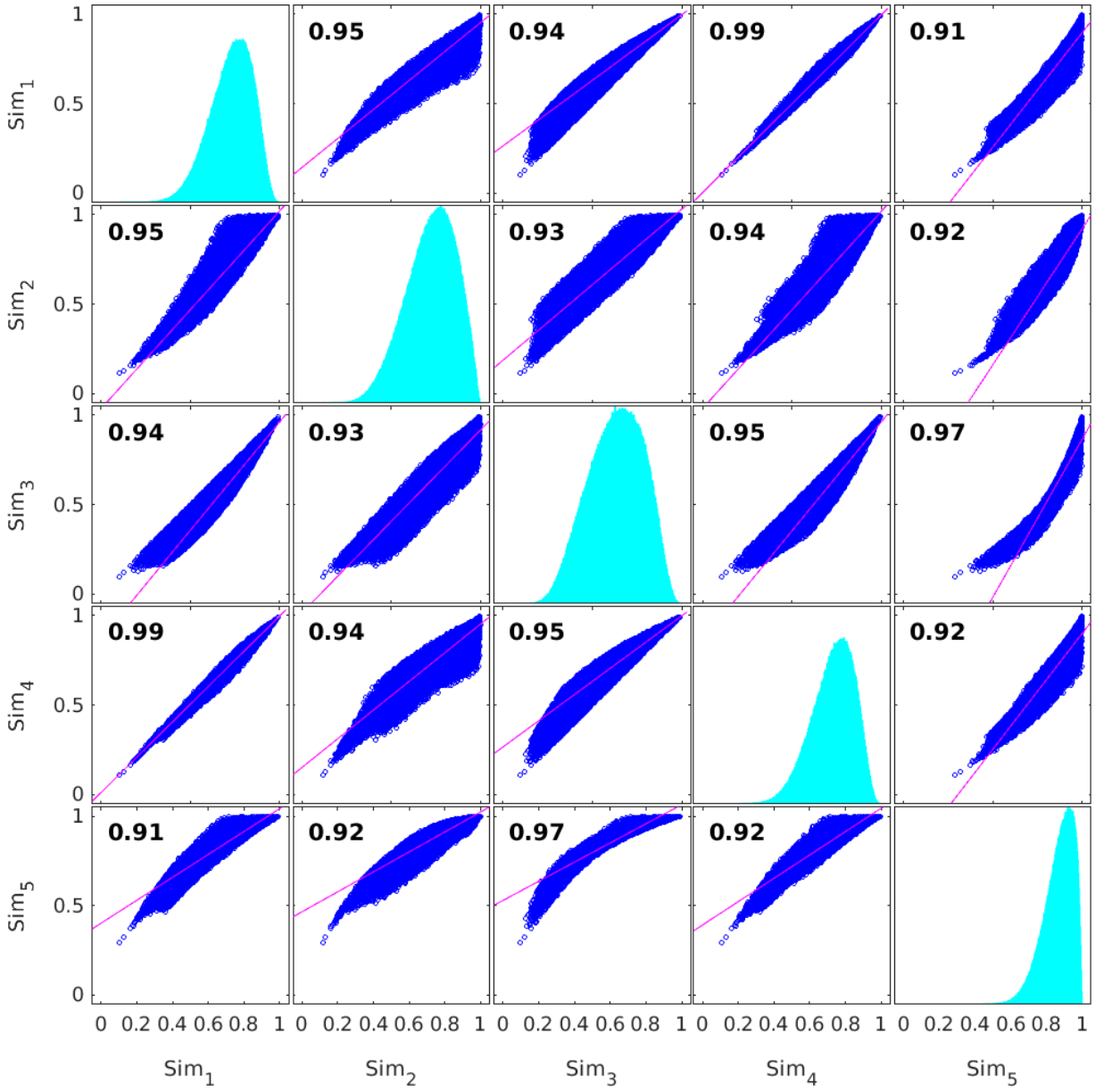


Figure 1 Scatter plots for each pair of similarity measures, each contains one million points with coordinates $[s_j(A_i, B_i), s_k(A_i, B_i)]$, $j, k = 1, \ldots, 5$, $j \neq k, i = 1, \ldots, 1000000$. Plots on the main diagonal represent histograms of the similarity values obtained for the one million pair of randomly generated trapezoidal fuzzy numbers.

to compensate for some discrimination inability does not change the nature of the original similarity measure significantly. We can see that the similarities (as desired by their authors) focus on slightly different aspects of the fuzzy numbers. Still, they remain quite interchangeable (the lowest correlation between the values of two similarities was found to be 0.91).

The findings presented in this paper complement the analyses of the performance of the similarity measures already proposed in the literature on specific examples by carrying out a large-scale numerical experiment. Even though the correlations between the similarity values are found to be rather high, there are differences, mainly in terms of the actual values of the similarities. It is for example "easier" to obtain a high similarity value using s_5 than s_3. Note, that the results presented in this paper are valid for trapezoidal fuzzy numbers with the height of one. For generalized trapezoidal fuzzy numbers the correlations are expected to be lower.

Acknowledgements

This research would like to acknowledge the funding received from the Finnish Strategic Research Council, grant number 313396 / MFG40 - Manufacturing 4.0 and by the Specific University Research Grant IGA_FF_ 2019_002, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2019.

References

[1] Chen, S.-J., and Chen, S.-M.: Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on Fuzzy Systems* **11** (2003), 45–56. URL http://ieeexplore.ieee.org/document/1178065/.

[2] Hejazi, S., Doostparast, A., and Hosseini, S.: An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **38** (2011), 9179–9185. URL http://linkinghub.elsevier.com/retrieve/pii/S0957417411001217.

[3] Khorshidi, H. A., and Nikfalazar, S.: An improved similarity measure for generalized fuzzy numbers and its application to fuzzy risk analysis. *Applied Soft Computing* **52** (2017), 478–486. URL http://linkinghub.elsevier.com/retrieve/pii/S1568494616305336.

[4] Klir, G. J., and Yuan, B.: *Fuzzy Sets and Fuzzy Logic: Theory and Aplications*. Prentice Hall, New Jersey, 1995.

[5] Patra, K., and Mondal, S. K.: Fuzzy risk analysis using area and height based similarity measure on generalized trapezoidal fuzzy numbers and its application. *Applied Soft Computing* **28** (2015), 276–284. URL https://linkinghub.elsevier.com/retrieve/pii/S1568494614006061.

[6] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016* (Slavíčková, P., ed.). Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[7] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* **29** (2017), 485–503.

[8] Talášek, T., and Stoklasa, J.: Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance / similarity affect the ordering. In: *Proceedings of the 36th International Conference on Mathematical Methods in Economics* (Váchová, L., and Kratochvíl, O., eds.). MatfyzPress, Jindřichův Hradec, 573–578.

[9] Wei, S.-H., and Chen, S.-M.: A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **36** (2009), 589–598. URL http://linkinghub.elsevier.com/retrieve/pii/S0957417407004630.

[10] Wen, J., Fan, X., Duanmu, D., and Yong, D.: A modified similarity measure of generalized fuzzy numbers. *Procedia Engineering* **15** (2011), 2773–2777. URL https://linkinghub.elsevier.com/retrieve/pii/S1877705811020236.

Business environment and its relations within Travel and Tourism Competitiveness Index

Eva Litavcová[1], Petra Vašaničová[2]

Abstract. Economic dataset concerning The Travel and Tourism Competitiveness Index (TTCI) published by World Economic Forum consists of 14 characteristics (pillars) of tourism competitiveness. Pillars, which are composed of 90 subpillars, are divided into 4 main subindexes. Business Environment (BE) pillar (consisted of 12 subpillars) is the part of Enabling Environment subindex. To increase the country's tourism competitiveness, state of the business environment is an important piece of information for potential investors. The aim of this paper is an investigation of relation of 13 other indices of the TTCI from year 2017 to Business Environment. We used regression quantiles and lasso estimates for modelling. In resulting table is shown, which of 13 pillars are in relation to Business Environment on different quantile levels. For the median of BE, the strongest relationship was proven with HRLM (Human Resources and Labour Market). Other statistically significant relationships were verified with other seven pillars, one of them is ATI (Air Transport Infrastructure) for example.

Keywords: The Travel and Tourism Competitiveness Index, Regression quantiles, Lasso estimates.

JEL Classification: C31, L83, Z32
AMS Classification: 62H20, 91B82

1 Introduction

To examine overall tourism competitiveness has a rapidly growing importance in the ever-evolving tourism industry. One of the widely used indicators of the travel and tourism (T&T) competitiveness of countries all over the world is Travel and Tourism Competitiveness Index (TTCI). This index provides a comprehensive strategic benchmarking tool and contributes to the development and competitiveness of a country. From a methodological point of view used for the period 2015-2017, the top indicator TTCI consists of four subindexes, which are created by several pillars and these pillars are made by further subpillars. For each country under consideration, the index is given by the overall score. Five pillars of Enabling Environment subimdex characterise the main settings required for operating in a country. Four pillars of subindex T&T Policy and Enabling Conditions characterise particular policies or strategic aspects that affect the T&T industry more directly. Three pillars of Infrastructure characterise the availability and quality of physical infrastructure in a country. Finally, two pillars of Natural and Cultural Resources identify the main reasons or motives to travel [4]. We are interested in a question, how the other 13 pillars of T&T competitiveness influence and relate to Business Environment. We aim: (A) to answer this question by using quantile regression and penalized lasso estimates, (B) to compare their results. The aim of this paper is an investigation of relation of 13 other indices of the TTCI from year 2017 to Business Environment (*BE*). *BE* is the first pillar of the subindex Enabling Environment that is directly linked to the economic growth of a country. *BE* pillar captures the extent to which a country has a supportive policy environment for businesses. Jin and Weber [10] suggest that general business environment is one of the major factors of tourism competitiveness for event organizers. At the beginning, we describe the importance of the other 13 pillars that make up the index [4]. Safety and Security (*SS*) is an essential pillar because tourists usually do not like to travel to dangerous countries and regions. Also healthy environments and safe landscapes, in terms of Health and Hygiene (*HH)*, can make a significant contribution to increasing the visitor arrivals into the destination. Moreover, low readiness of human resources can be a barrier to the tourism industry development. Next, within TTCI, Human Resources and Labour Market (*HRL*) pillar is evaluated. Accelerating progress in technology related to the Internet has led to progressive changes [19]. Therefore, ICT Readiness (*ICT*) of the country is also a very important pillar describing the competitiveness of the destination. Indicators represent Prioritization of T&T (*TT*) are part of sixth pillar of TTCI. Country openness in terms of travel, visa conditions and other restrictive policies are part

[1] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, eva.litavcova@unipo.sk.
[2] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, petra.vasanicova@smail.unipo.sk.

of the International Openness (*IO*). Pricing issue in market conditions is also the specific part of TTCI and compose Price Competitiveness (*PC*) pillar. The issues of environmental sustainability in relation to tourism are part of Environmental Sustainability (*ES*) pillar. Air Transport Infrastructure (*ATI*), Ground and Port Infrastructure (*GPI*) and Tourist Service Infrastructure (*TSI*) are pillars represent infrastructure. Unique and distinct natural environment provide resources for tourism development, therefore Natural Resources (*NR*) contributes to tourism competitiveness. Finally, culture is one of the most significant tourist's motivators in choosing a given destination [3], and thus the Cultural Resources and Business Travel (*CRBT*) pillar is one of the very important components of the TTCI.

2 Data and Methodology

Our dataset consist of values of Travel and Tourism Competitiveness Index (TTCI) and its pillars among 136 economies from 2017. The question is how the other 13 pillars of TTCI influence and relate to Business Environment. This article aims to answer this question by using quantile regression and penalized lasso estimates. The effect of a change in the Business Environment pillar of the country among 136 countries depending on other 13 pillars contributing to TTCI would be tested by using simple ordinary least squares method (OLS), which describes conditional mean of response variable as a linear function of the explanatory variable. The response and all the regressors are continuous variables without missing values. However, 1-2 outliers occur in 8 variables and in one up to 10. And moreover, the collinearity diagnostics confirm that there are serious problems with multicollinearity. Several eigenvalues are close to 0, indicating that the predictors are highly intercorrelated and that small changes in the data values may lead to large changes in the estimates of the coefficients. The condition indices are computed as the square roots of the ratios of the largest eigenvalue of X^TX to each successive eigenvalue. Six of these indices are larger than 30 (from 31.2 to 78.9), suggesting a serious problem with collinearity [9], [17]. Thus, other methods should be used (see [11], [12]). In this study, we focus on mentioned relationships of investigated variables by using (1) quantile regression and by using (2) lasso estimates. According to Agresti [1], quantile regression models quantiles of a response variable as a function of explanatory variables. This method can be less severely affected by outliers than OLS. However, when the normal linear model truly holds, the least squares estimators are much more efficient. The OLS model estimates constant effects of the independent variables on the conditional mean of the dependent variable and assumes a normal distribution of errors with constant variance. Quantile regression models were initially proposed by Koenker and Bassett [13] as a method of robust regression that would account for a non-normal distribution of error terms and as a test for heteroskedastic error terms (see more in Koenker [15] and [16]). At the heart of the optimization problem, there is the minimization of asymmetrically weighted absolute residuals. By asymmetrically weighted residuals, they mean assigning different weights to positive and negative residuals [2]. This results in the minimization equation for conditional quantiles,

$$\min_{\xi} \sum_{i=1}^{n} \rho_{\tau}(y_i - \xi(x, \beta)) \tag{1}$$

where $\rho(.)$ is the absolute value function that gives the τ th sample quantile, y_i is the observed value of the dependent variable, and $\xi(x, \beta)$ is the predicted value in the form of parametric function [2]. Koenker [15, p. 10] specify the τ-th regression quantile function as $Q_Y(\tau|x) = x^T\beta(\tau)$, and consider of $\hat{\beta}(\tau)$ solving

$$\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i^T\beta) \tag{2}$$

with check function $\rho_\tau(u) = u \cdot \tau$ for $u \geq 0$ and $\rho_\tau(u) = u \cdot (\tau - 1)$ for $u < 0$ (and with respect to the design matrix described below). Quantile regression problem may be reformulated as a linear program [15].

As reported by Fenske [5, p. 7-8], quantiles are defined based on the cumulative distribution function (*cdf*) F_Y of a continuous random variable *Y*. The $\tau \,.\, 100\%$ quantile of *Y* can be written as a value y_τ, where

$$F_Y(y_\tau) = P(Y \leq y_\tau) = \int_{-\infty}^{y_\tau} f(u)du = \tau \tag{3}$$

for $\tau \in (0,1)$. It is only unique if F_Y is strictly monotonic increasing. In case that information on an additional random variable *X* is given, the quantile can similarly be expressed conditional on *X* is equal *x*:

$$F_Y(y_\tau(x)|\, X = x) = P(Y \leq y_\tau(x)|\, X = x) = \tau. \tag{4}$$

The quantile function $Q_Y(\tau|X = x)$ is defined as the smallest *y* where the quantile property is fulfilled. If F_Y is strictly increasing, the quantile function is set to the inverse of the *cdf* of *Y*. The relationship between quantile function and *cdf* can be expressed as

$$F_Y(y_\tau(x)|\, X = x) = \tau \;\Leftrightarrow\; Q_Y(\tau|\, X = x) = y_\tau(x) \tag{5}$$

for strictly increasing F_Y, which emphasizes that the quantile function describes $\tau \,.\, 100\%$ quantiles of *Y* depending on covariates *x* and a quantile parameter $\tau \in (0,1)$.

Quantile regression is an approach to model the conditional quantile function of a continuous variable of interest *Y*, e.g. response variable, depending on further variables or covariates *X*. In the linear model it can be expressed as $y_i = x_i^T \beta_\tau + \varepsilon_{\tau i}$, $i = 1, \dots, n$. The index $i = 1, \dots, n$, denotes the observation, y_i is the response value and $x_i = (1, x_{i1}, \dots, x_{ip})^T$ the given covariate vector for observation i. The quantile-specific linear effects are denoted by $\beta_\tau = (\beta_{\tau 0}, \beta_{\tau 1}, \dots, \beta_{\tau p})^T$, and $\tau \in (0,1)$ indicates a quantile parameter which has to be fixed in advance. The random variable $\varepsilon_{\tau i}$ is assumed to be an unknown error term with *cdf* $F_{\varepsilon_{\tau i}}$ and density $f_{\varepsilon_{\tau i}}$ depending on quantile parameter τ and observation i. For quantile regression, no specific assumptions are made apart from $\varepsilon_{\tau i}$ and $\varepsilon_{\tau j}$ being independent for $i \neq j$, and $\int_{-\infty}^{0} f_{\varepsilon_{\tau i}}(\varepsilon_{\tau i})\, d\varepsilon_{\tau i} = F_{\varepsilon_{\tau i}}(0) = \tau$. Due to this assumption, the quantile function $Q_{Y_i}(\tau|x_i)$ of the response variable Y_i conditional on covariate vector x_i at a given quantile parameter τ is equal to $x_i^T \beta_\tau$. Thus, the parameter $\beta_{\tau 1}$, for example, can be interpreted as the change of the conditional quantile function when x_{i1} changes to $x_{i1} + 1$, given all other covariates remain constant [5].

The lasso is a popular method for regression that uses an l_1 penalty to achieve a sparse solution. Regression shrinkage and selection via the lasso was proposed by Tibshirani in 1996. As states in his work [20, p 267], the lasso minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. It produces interpretable models like subset selection and exhibits the stability of ridge regression. The lasso does not focus on subsets but rather defines continuous shrinking operation that can produce coefficients that are exactly 0. For standardized predictor variables $x_1, x_2, \dots, x_p$ (assumed independent *n* observations for $i = 1,2, \dots, n$) the lasso estimate is defined by [20, p 268]

$$(\hat{\beta}_0, \hat{\beta}) = \operatorname{argmin}\left\{\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_j \beta_j x_{ij}\right)^2\right\} \quad \text{subject to} \quad \sum_j |\beta_j| \leq t \tag{6}$$

where $\hat{\beta} = \left(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p\right)^T$. For all *t*, the solution for β_0 is $\hat{\beta}_0 = \bar{y}$. Here $t \geq 0$ is a tuning parameter which controls the amount of shrinkage that is applied to the estimates. The prediction error (mean square error of an estimate $\hat{Y}$ plus variance of residuals) is estimated over a grid of values of *s* (*s* is normalized lasso parameter $s = \text{ter}\; s = t/\sum|\hat{\beta}_j^0|$, where $\hat{\beta}_j^0$ are the full least squares estimates) from 0 to 1 inclusive. The value $\hat{s}$ yielding the lowest prediction error is selected. Cross-validation and other two methods are employed for the estimation of the lasso parameter *t* (more in [20]).

Moreover, in 2008 [6] was proposed fast algorithms for fitting generalized linear models with elastic-net penalties. Elastic net penalty is a mixture of the l_1 (lasso) and l_2 (ridge regression) penalties [18]. The algorithm for the elastic net includes the lasso and ridge regression as special cases ([7], [8]). The elastic net solves the following problem (from [6, p. 3], more general see in [8])

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \left[\frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right]\right] \tag{7}$$

over a grid of values of λ covering the entire range. The tuning (regularization) parameter λ controls the overall strength of the penalty. The second sum in (7) is the *elastic-net penalty*. For $\alpha = 0$ it is the ridge regression, and for $\alpha = 1$ the lasso penalty. The elastic net with $\alpha = 1 - \varepsilon$ for some small $\varepsilon > 0$ performs much like the lasso, but removes any degeneracies and wild behavior caused by extreme correlations [6].

3 Results

Table 1 shows the coefficients of the *BE* as dependent variable, which were estimated at the 15th, 25th, 40th, 50th, 60th, 75th and 85th quantile levels using multiple quantile regression on the 13 pillars contributing to TTCI. The tests of significance were based on the robust bootstrap estimations of standard errors [14], [16]. There can be seen in detail the relationship of each selected quantiles of the dependent variable on the chosen five variables. Three out of four covariates having no significant coefficient were omitted from the table (*IO*, *ES*, and *TSI*). To achieve the interpretation of the results presented in Table 1, a compact way of presenting information on the OLS and the quantile regression coefficients is with a series of effects displays in Figure 1. In each panel, the value of the OLS coefficient is represented by a solid line, the values representing a 95% confidence interval

around the OLS coefficient are represented with dashed lines. The quantile regression coefficients are represented by the black dots, and a 95% confidence band around the quantile regression coefficient estimates is represented by the grey band.

tau	SS	HH	HRL	ICT	TT	PC	ATI	GPI	NR	CRBT
0,15	0.025	-0.426[c]	0.659[b]	0.351	0.094	0.228	0.110	0.120	-0.140	-0.058
0,25	0.069	-0.390[d]	0.513[b]	0.314[b]	0.043	0.223[b]	0.132	0.114	-0.105	-0.087
0,40	0.014	-0.300[c]	0.377[b]	0.296[b]	0.112	0.143	0.145	0.165[a]	-0.071	-0.080
0,50	0.037	-0.189[b]	0.351[b]	0.241[b]	0.162[b]	0.003	0.230[c]	0.137[a]	-0.049	-0.128[b]
0,60	0.051	-0.231[d]	0.383[c]	0.222[b]	0.138[b]	0.031	0.246[c]	0.093	-0.067	-0.111[a]
0,75	0.095	-0.181[c]	0.431[d]	0.154	0.066	0.047	0.111	0.152[b]	-0.042	-0.065
0,85	0.070	-0.191[c]	0.499[c]	0.234	0.004	0.032	0.100	0.167[a]	-0.039	-0.088
OLS	0.044	-0.269[d]	0.515[d]	0.219[c]	0.115[b]	0.090	0.168[c]	0.136[b]	-0.111[b]	-0.073[b]

Table 1 Estimated coefficients (calculated in *R*, employed package *quantreg* [16]) of Business Environment as dependent variable - a results of multiple quantile regression; "*a*", "*b*", "*c*", "*d*" denotes 10%, 5%, 1%, and 0.1% significance level.

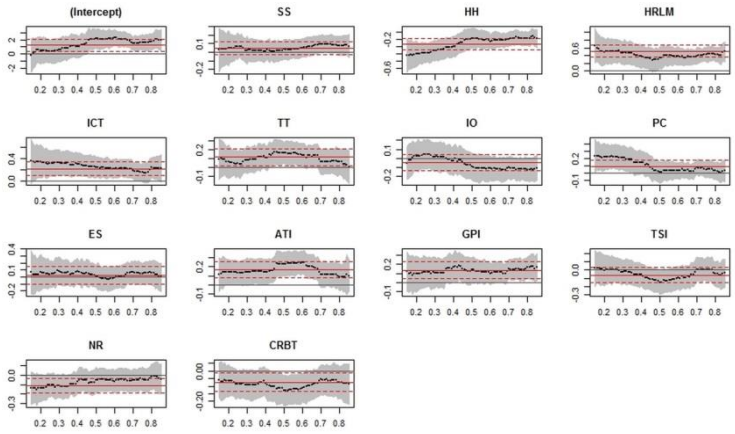


Figure 1 Quantile regression process for Business Environment (calculated in *R*, employed package *quantreg*)

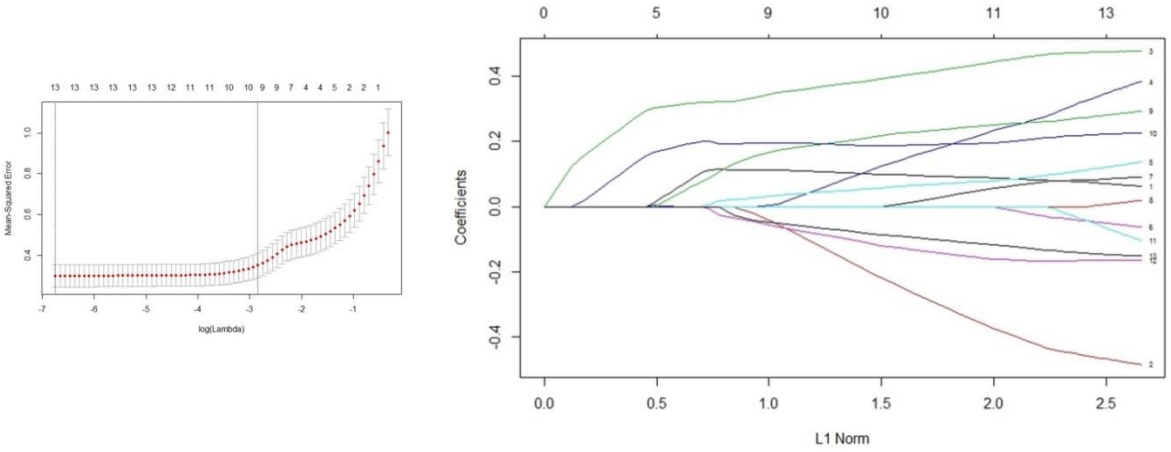


Figure 2 (calculated in *R*, employed package *glmnet* [8]) left: Lasso - the cross-validated error rates; right: Lasso estimates of regression parameters for all 13 regressors; coefficients are numbered as follows: 1-*SS*, 2-*HH*, 3-*HRLM*, 4-*ICT*, 5-*TT*, 6-*IO*, 7-*PC*, 8-*ES*, 9-*ATI*, 10-*GPI*, 11-*TSI*, 12-*NR*, 13-*CRBT*.

The quantile regression effect of *HH* reveals a pattern of increasing effect, but it is never statistically different from the value of the OLS coefficient. After exhibiting a nearly linear increase in value through the median, the value of the quantile regression coefficient are almost constant on the upper bound of the OLS estimate. An interesting phenomenon is seen taking into account regressor *NR*, at which no significance at any quantile level has been demonstrated despite significant OLS coefficient. Regressor *PC*, which was not significant in OLS, has

significant coefficients at low quantile levels for tau from 0.15 to 0.35, while at upper quantile levels it has not. For the regressors *TT*, *ATI*, *GPI*, and *CRBT*, the wavy shape of the coefficient curve depending on quantile level signify their significance only at certain intervals of quantile levels.

When considering the lasso estimates for our variables, which are there standardized, Figure 2 (left) plots of the cross-validated error rates. Each dot represents a λ value along our path, with error bars to give a confidence interval for the cross-validated error rate. The left vertical bar indicates the minimum error (denoted *lambda.min*), while the right shows the largest value of λ such that the error is within one standard deviation of the minimum (denoted *lambda.1se*). The top of the plot gives the size of each model [6]. If right vertical bar will be considered, 9 regressors stay in the final model. It is clear from the Figure 2 (right), which variables there are (from 13 variables). If the comparable lasso coefficients of the standardized predictors for *lambda.1se* are ranked according to size in descending order, their ranking is as follows: *HRLM*, *HH*, *ATI*, *GPI*, *ICT*, *NR*, *SS*, *CRBT*, *TT*, *PC*, and for *IO*, *ES*, *TSI* are zeros.

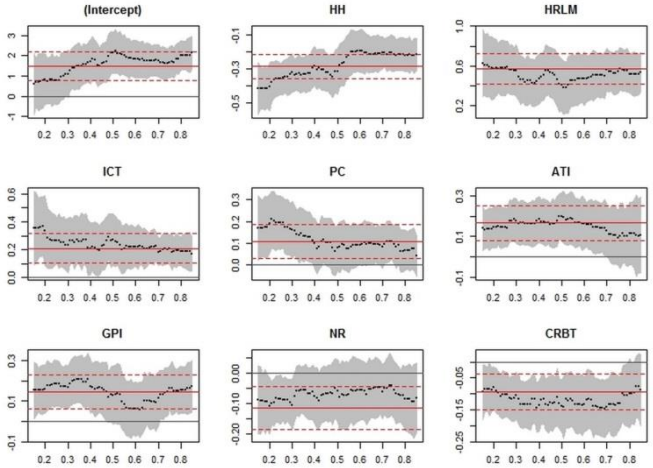


Figure 3 Quantile regression process for Business Environment with 8 regressors (calculated in *R*, lib. *quantreg*)

For OLS regression, with all 13 regressors for dependent *BE*, was used stepwise procedure that eliminated 5 variables: *SS*, *TT*, *IO*, *ES*, and *TSI*. Result of quantile regression using 8 remained regressors is presented in Figure 3. By comparing Figure 1 and Figure 3, it is possible to find out that the pattern of development of quantile coefficient in Figure 3 is very similar to Figure 1, taking into account 8 selected regressors; also in relation to confidence bounds of OLSs. Taking into account regressors *PC*, *ATI*, *GPI*, *NR*, and *CRBT*, we can see more extensive sections of quantile levels, in which the corresponding coefficient is significantly different from 0.

4 Conclusion

Our goal was to reveal how the other 13 indices (pillars) concerning competitiveness in tourism influence and relate to the pillar Busines Environment (*BE*). For the first, standard OLS and quantile regression for all 13 regressors on the *BE* as dependent variable was used. Second, the lasso estimates was performed. Third, on set of regressors after elimination by stepwise OLS procedure quantile regression was used.

Our quantile regression results finally show a stable positive relationship between *BE* and *HRLM*. Next shows that a negative relationship between *BE* and *HH* is more and more weaker with increasing quantile level of *BE*. Similarly, but very softly, it is within a positive relationship between *BE* and *ICT*, *PC* and *ATI*. However, taking into account *PC* and *ATI* on higher quantile levels, we can see no significant relationship with *BE*. *NR* coefficients are significantly different from 0 only in a narrow interval around the quantile level 0.25. *GPI* coefficients are not significantly different from 0 in interval circa from 0.45 to 0.7. *CRBT* coefficients are significantly different from 0 up to quantile level around 0.8.

When using LS-lasso regression on standardized and strong correlated variables, it is possible to see some deviation from the output of quantile regressions performed on both the original and reduced regression sets. The value of *lambda.min*, e.g. lambda that gives minimum mean cross-validated error was obtained in 0.00205 with df=13, thus all of 13 variables are taking into the model. If we took into account *lambda.1se*, the largest value of lambda such that error is within 1 standard error of the minimum, λ=0.04846, the 9 regressors stay. In 3 cases, these are different from the selection of OLS stepwise. First, using lasso we have included variables *SS* and *TT*,

and to the contrary, in the selection we miss *PC* variable. Experimenting with parameter α adjustment to a value slightly less than 1 for elastic net there was no significant change of df.

In the next step it is needed to use stepwise RQ procedure, RQ-lasso, and continue experimenting with the elastic net. The data set used here is useful in experimenting with different regression approaches taking into account their assumptions. However, there is a problem in a construction of variables that are not natural but their construction is based on subjective considerations and managerial ideas what result in complex indicators. Therefore, for next research, we would recommend making some investigation into which indices can be related and then support or decline such hypotheses.

Acknowledgements

This contribution was supported by research grant VEGA No. 1/0470/18 "Economic activity of tourism in the European area", and GAPU 43/2019 "Cultural differences and their impact on tourism". This work was supported by the Slovak Research and Development Agency under the contract no. APVV-17-0166.

References

[1] Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, New Jersey: John Wiley & Sons.
[2] Britt, C. L. (2009). Modeling the distribution of sentence length decisions under a guidelines system: An application of quantile regression models. *Journal of Quantitative Criminology*, 25(4), 341–370.
[3] Correia, A., Kozak, M. & Ferradeira, J. (2013). From tourist motivations to tourist satisfaction. *International Journal of Culture, Tourism and Hospitality Research*, 7(4), 411–424.
[4] Crotti, R. & Misrahi, T. (2017). *The Travel & Tourism Competitiveness Report 2017: Paving the way for a more sustainable and inclusive future*. Geneva: The World Economic Forum.
[5] Fenske, N. (2012). *Structured additive quantile regression with applications to modelling undernutrition and obesity of children*. Dissertation thesis. Munchen.
[6] Friedman, J., Hastie, T. & Tibshirani, R. (2008). Regularization Paths for Generalized Linear Models via Coordinate Descent, https://web.stanford.edu/~hastie/Papers/glmnet.pdf. *Journal of Statistical Software*, Vol. 33(1), 1-22, http://www.jstatsoft.org/v33/i01/
[7] Friedman, J., Hastie, T., Tibshirani, R., Simon, N., Narasimhan, B. & Qian, J. (2018). *Package 'glmnet', April 2, 2018*. https://cran.r-project.org/web/packages/glmnet/glmnet.pdf
[8] Hastie, T. & Qian, J. (2016). *Glmnet Vignette*. https://cran.project.org/web/packages/glmnet/vignettes/glmnet_beta.pdf
[9] IBM SPSS help pages. *.../help/index.jsp?topic=/com.ibm.spss.statistics.cs/casestudies_intro.htm*
[10] Jin, X. & Weber, K. (2016). Exhibition destination attractiveness–organizers' and visitors' perspectives. *International Journal of Contemporary Hospitality Management*, 28(12), 2795–2819.
[11] Jurczyk, T. (2012). Outlier Detection under Multicollinearity. *Journal of Statistical Computation and Simulation*, 82 (2), 261-278.
[12] Kalina, J. & Peštová, B. (2017). Exact Inference in Robust Econometrics under Heteroscedasticity. *Proceedings of 11th International Days of Statistics and Economics MSED 2017* (pp. 636-645), Prague, Czech Republic.
[13] Koenker, R. & Bassett, G. (1978). 'Regression Quantiles'. *Econometrica*, 46 (1), 33-50.
[14] Koenker, R. & Hallock, K. F. (2001). 'Quantile regression'. *The Journal of Economic Perspectives*, 15 (4), 43-56.
[15] Koenker, R. (2005). *Quantile regression*. Cambridge University Press, New York
[16] Koenker, R. (2018). *Quantile regression in R: a vignette*. https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf
[17] Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool, Second Edition*. New York: Springer-Verlag, Inc.
[18] Simon, N., Friedman, J., Hastie, T., Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, 39(5), 1-13, http://www.jstatsoft.org/v39/i05/
[19] Štefko, R., Fedorko, I., Bačík, R. & Fedorko, R. (2015). An analysis of perceived topicality of website content influence in terms of reputation management. *Polish Journal of Management Studies*, 12(1), 177–185.
[20] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288, http:/www.jstor.org

Price Clustering Phenomenon

Petra Tomanová[1]

Abstract. This paper studies the price clustering phenomenon which refers to an excessive occurrence of transaction prices at certain fractions or digits. The source of the phenomenon are traders who execute the transaction in multiples of dimes (10 cents – the second digit is 0) or nickels (5 cents – the second digit is 0 or 5). First, the paper summarizes the state of the art on the price clustering phenomenon from its first observation at the NYSE in 1962 by M. F. M. Osborne, through the year 1994 when price clustering phenomenon has drawn much attention due to the study of NASDAQ stocks of Christie and Schultz, till the most recent papers on clustering phenomenon. Second, the paper contributes to the knowledge by investigating the (conditional) probabilities of prices which evinces bumpy behavior due to the price clustering phenomenon. Our findings are useful in parametric modeling of high-frequency prices with discrete values. In empirical study, we focus on stocks traded on NYSE and NASDAQ exchanges.

Keywords: NASDAQ, NYSE, price clustering, price modeling.

JEL Classification: G14
AMS Classification: 62-07

1 Price clustering

In an efficient market, the last digits of the price should exhibit a uniform distribution. When some digits occur more frequently than others, it is referred to as price clustering. This paper studies the excess occurrence of zeros and fives in terms of decimal places of precision, data granularity and conditional probabilities of their occurrence. The source of the phenomenon are traders who execute transactions in multiples of dimes (10 cents – the second digit is 0) or nickels (5 cents – the second digit is 0 or 5).

The first academic paper on the price clustering is dated to 1962, when Osborne [7] rigorously described clustering by conducting an empirical study of US stock prices. Using the words of Osborne [7]: *"... there is a pronounced tendency for prices to cluster on whole numbers, halves, quarters, and odd one-eighths in descending preference, like the markings on a ruler"*. Harris [4] added that clustering is pervasive and increases with price level and volatility, and decreases with capitalization and transaction frequency. The paper is followed by the work of Christie and Schultz [3] which has attracted much more attention to the price clustering phenomenon due to the empirical study of NASDAQ stocks.

Recent literature claims that price clustering is still remarkably persistent through time and across stocks, no matter if traded on developed or emerging markets. In the most recent papers, Lien et al. [5] studied the order price clustering, size clustering, and stock price movements in an emerging market, the Taiwan Stock Exchange (TWSE); and Mishra and Tripathy [6] investigated the price and trade clustering in an emerging market using the NSE listed stocks.

Moreover, Mishra and Tripathy [6] pointed out that the most of previous studies on price clustering have focused on quote-driven markets and less on order-driven markets. Our empirical study focuses on the New York Stock Exchange (NYSE) and National Association of Securities Dealers Automated Quotations (NASDAQ) which are considered as hybrid markets. Hybrid markets combine attributes both from the quote- and order-driven systems, e.g. those markets show the current bid and ask prices of the market makers and also allow traders to view all limit orders in the market.

2 Evidence from NYSE and NASDAQ

Our empirical study focuses on stocks traded on NYSE and NASDAQ exchanges. Eight most traded stocks (in terms of executed transactions) of Dow Jones Industrial Average (DJIA) index are analyzed. Specifically, four

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic, petra.tomanova@vse.cz

stocks traded on NYSE: ExxonMobil (XOM), General Electric[1] (GE), JPMorgan Chase (JPM) and Pfizer (PFE); and four stocks traded on NASDAQ: Apple (AAPL), Cisco Systems (CSCO), Intel (INTC) and Microsoft (MSFT) are selected. The study covers a period spanning from January 2 to May 31, 2018. Transaction data are extracted from the NYSE TAQ database.

2.1 Data issues

The high-frequency analysis on raw data extracted from the database might be misleading since the raw data includes trades with abnormal sale conditions; corrected, changed, cancelled or finished with error; non-standard trades such as warrant; and trades executed outside the trading hours. Thus, we performed the same high-frequency data cleaning as indicated in [2] which is a slightly adjusted version of the original data cleaning procedure described in [1].

The data are supposed to be stored with 2 decimal places of precision. However, some discrepancies and issues in raw data are present. Let the d-th decimal digit of price x be denoted as $x^{[d]}$. Table 1 reports the number of observations (as well as the percentage of cases) for which the $x^{[3]} > 0$, i.e. the data are stored with more than 2 decimal places of precision. Moreover, Table 1 reports a number of cases when digits on 3rd decimal place are equal to 5. When the discrepancy of decimal places of precision occurs, it is very likely that the 3rd decimal digit is equal to 5. Table 1 shows that this kind of data issue is negligible for AAPL and rather minor for the rest of stocks except for GE and PFE. In addition, Table 1 reports main data characteristic: sample mean of prices (mean), sample variance of prices (Var) and number of observations (nObs). Numbers of analyzed transactions (observations) range from 874,999 for PFE to 4,758,368 for AAPL.

	mean	Var	nObs	$x^{[3]} > 0$		$x^{[3]} = 5$	
AAPL	172.569	66.463	4,758,368	64,004	1.35%	64,004	1.35%
CSCO	42.868	3.788	1,903,507	79,732	4.19%	79,732	4.19%
INTC	49.586	12.605	2,505,538	87,275	3.48%	87,275	3.48%
MSFT	92.640	9.742	3,773,433	92,677	2.46%	92,677	2.46%
GE	15.126	2.516	1,192,833	651,797	54.64%	565,687	47.42%
JPM	111.945	9.085	1,685,676	83,981	4.98%	59,435	3.53%
PFE	35.902	1.154	874,999	220,864	25.24%	211,583	24.18%
XOM	78.841	19.2677	1,228,691	99,740	8.12%	85,906	6.99%

Table 1 Data characteristics

The *wrong decimal places of precision* (WDPP) issue is further analyzed in Figure 1 with respect to price clustering phenomenon. Figure 1 depicts histograms of digits for AAPL as a representative stock with a negligible WDPP issue and for GE and PFE as representative stocks with a serious WDPP issue. The issue is shown by histograms in the first column headed as *"3 decimal places: . . . "*.

Further, the WDPP issue is treated in two ways: (i) the 3rd decimal place is truncated, (ii) prices are rounded to 2 decimal places. Histograms in the second and third column of Figure 1 show the excess occurrence of individual digits in the second decimal place. Note that the second decimals should resemble the uniform distribution, i.e. each digit should occur with probability 10%. However, in AAPL case, we can see that the 0 and 5 digits are more frequent which we refer to as a price clustering phenomenon. Precisely, zeros occur in 14.667% of cases (excess occurrence is equal to 4.667%) when the 3rd decimal place is truncated and in 15.034% of cases (excess occurrence is equal to 5.034%) when the prices are rounded to 2 decimal places. Tables 2 and 3 report the percentage occurrence for all analyzed stocks when the 3rd decimal place is truncated and the prices are rounded to 2 decimal places respectively. For stocks with minor WDPP issue, the impact of different treatments (truncation vs. rounding) is negligible and the price clustering is clearly present.

However, when the WDPP issue is more profound, simply truncating the 3rd decimal place does not seem to be a correct way of dealing with this data issue. Histogram labeled as *"2 decimal places (trunc): PFE"* shows a distinct excess occurrence of 6 and 9 digits instead of the well-documented excess occurrence of 5 and 0 digits in the literature. It seems that for example the numbers such as $x^{[2:3]} = 45$ and $x^{[2:3]} = 95$ are actually distorted $x^{[2:3]} = 50$ and $x^{[2:3]} = 00$ respectively since when rounding is applied instead of truncation, the histogram *"2*

[1] General Electric was part of DJIA index from November 1907 till June 2018 which makes General Electric be the longest continuously present stock on the index. In June 2018, General Electric was replaced by Walgreens Boots Alliance.

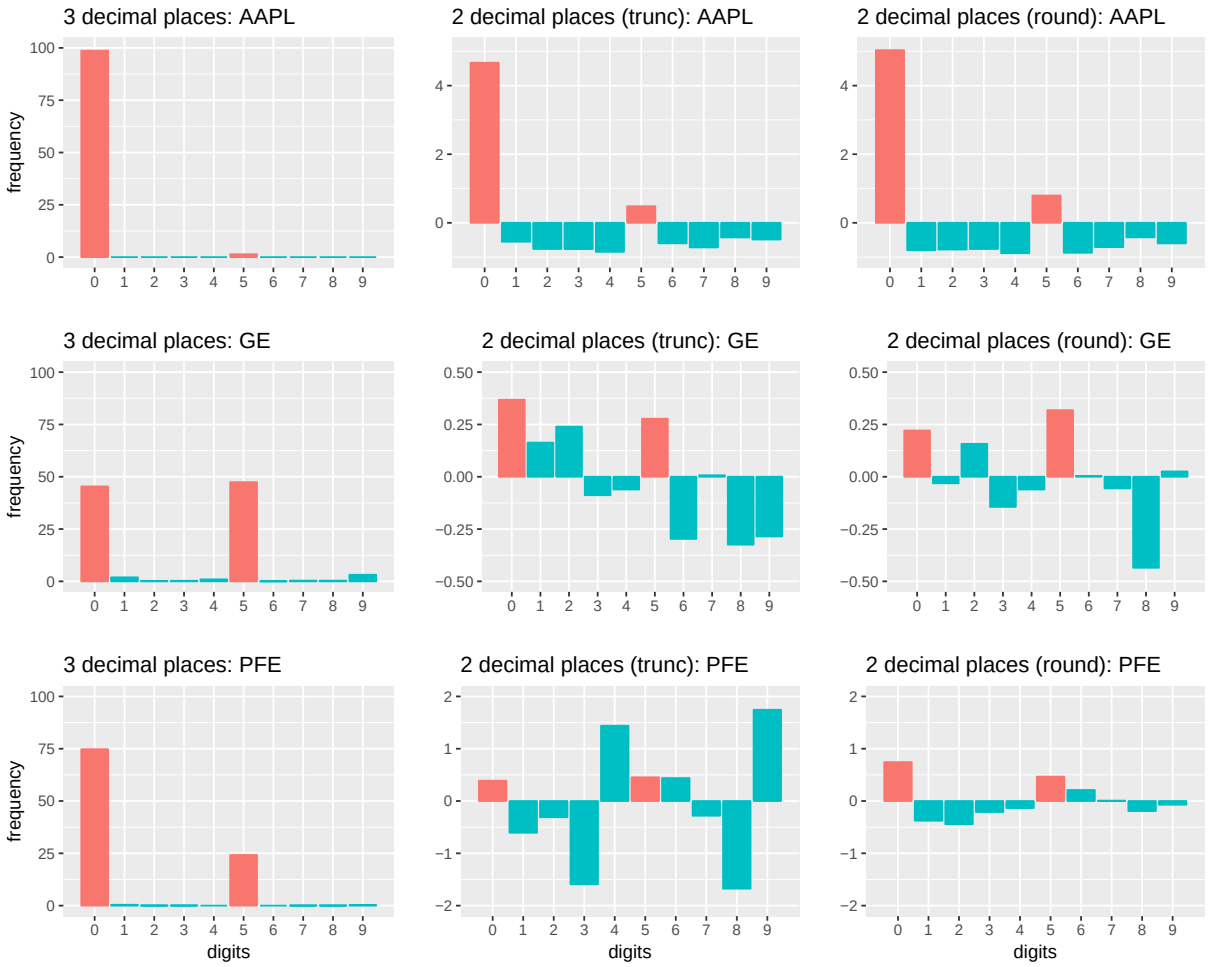


Figure 1 Percentage occurrence (1st column) and excess percentage occurrence (2nd and 3rd column) of digits

	d0	d1	d2	d3	d4	d5	d6	d7	d8	d9
AAPL	14.667	9.449	9.238	9.235	9.155	10.484	9.402	9.284	9.574	9.514
CSCO	11.240	9.629	9.663	9.263	10.126	10.329	9.859	9.809	9.624	10.459
INTC	12.066	9.742	9.636	9.616	9.685	10.496	9.513	9.555	9.831	9.862
MSFT	12.617	9.555	9.517	9.511	9.562	10.660	9.575	9.639	9.663	9.701
GE	10.368	10.163	10.240	9.912	9.939	10.277	9.703	10.007	9.676	9.715
JPM	11.086	9.886	9.848	9.743	9.741	10.126	9.696	9.807	10.035	10.031
PFE	10.388	9.394	9.690	8.410	11.439	10.453	10.438	9.719	8.325	11.744
XOM	10.745	9.758	8.504	11.333	10.076	10.288	9.877	8.349	11.211	9.859

Table 2 Percentage occurrence of second decimals when 3rd decimal place is truncated

decimal places (round): PFE" is similar to histograms of "healthy" stocks, i.e. without a significant WDPP issue. In spite of the fact that it is still not fully clear what the source of the issue is, this finding might help to treat the issue correctly.

	d0	d1	d2	d3	d4	d5	d6	d7	d8	d9
AAPL	15.034	9.202	9.222	9.237	9.114	10.795	9.127	9.292	9.577	9.402
CSCO	11.200	9.720	9.676	9.603	9.730	10.371	9.771	9.926	9.934	10.070
INTC	12.060	9.771	9.638	9.621	9.679	10.496	9.517	9.550	9.811	9.858
MSFT	12.571	9.623	9.509	9.507	9.556	10.602	9.629	9.660	9.658	9.685
GE	10.221	9.969	10.157	9.856	9.939	10.318	10.004	9.945	9.565	10.025
JPM	11.034	9.976	9.850	9.737	9.741	10.069	9.758	9.797	10.026	10.012
PFE	10.744	9.624	9.557	9.790	9.866	10.466	10.211	10.007	9.808	9.929
XOM	10.710	9.828	9.921	9.977	9.944	10.262	9.904	9.811	9.856	9.787

Table 3 Percentage occurrence of second decimals when prices are rounded to 2 decimal places

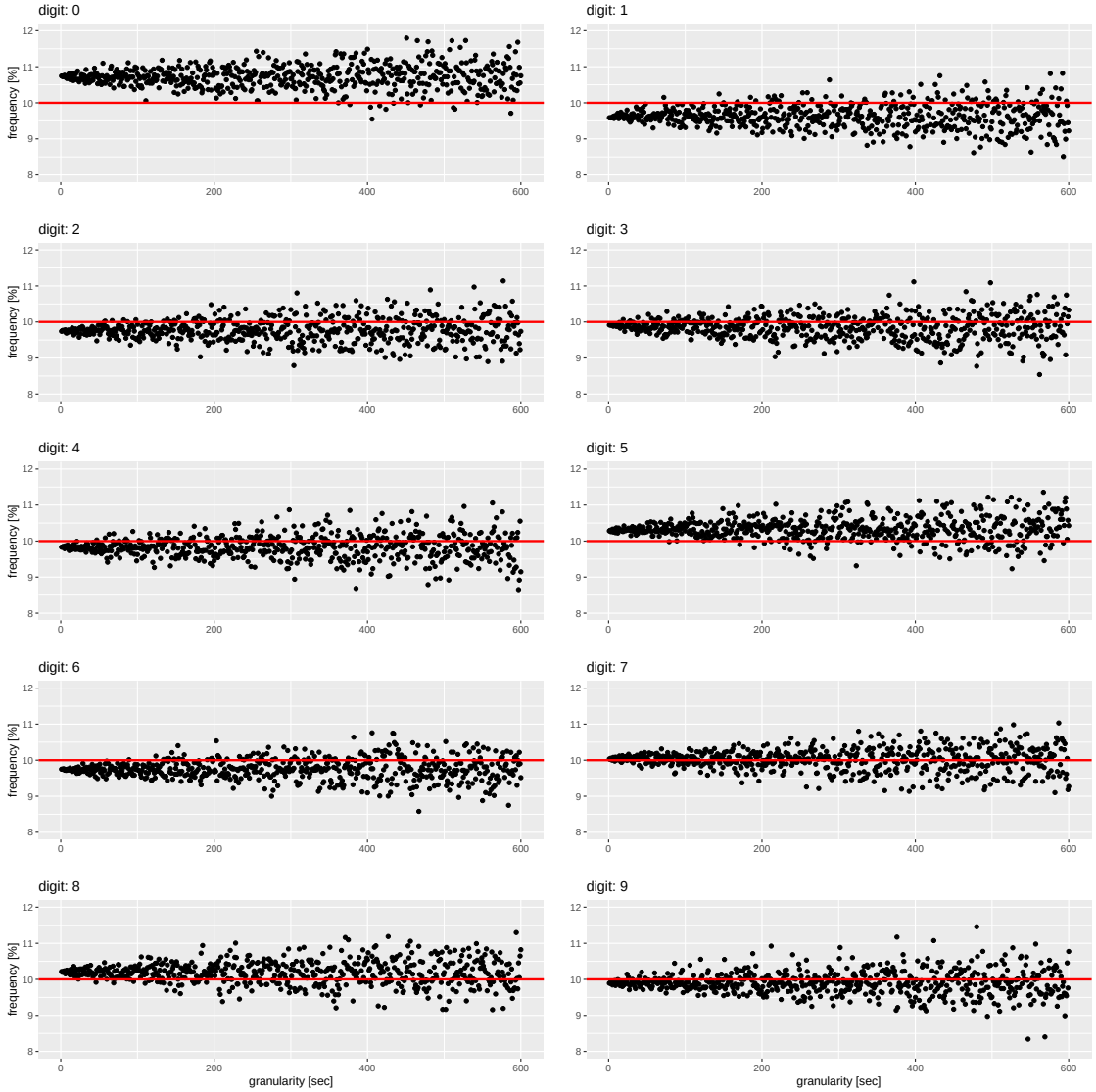


Figure 2 Percentage occurrence of digits based on different data granularity for AAPL stock

2.2 Granularity

Until now, the tick data were analyzed and the presence of price clustering phenomenon was confirmed based on the tick data. In this section, the price clustering is investigated in terms of different data granularity. The goal is to find out whether the price clustering is persistent when tick data are aggregated to second or minute data. Figure 2 shows results for AAPL prices when second decimals are analyzed. The percentage occurrence of digits is plotted on y-axis and time aggregation on k seconds, $k = 1, \ldots, 600$, is on x-axis. Red line represents the theoretical percentage when the second decimals resemble the uniform distribution. We can observe that the clear pattern of price clustering is present in the data even when aggregating data up to 10 minutes since the vast majority of observations (black dots) lie above the theoretical percentage for 0 and 5 digits.

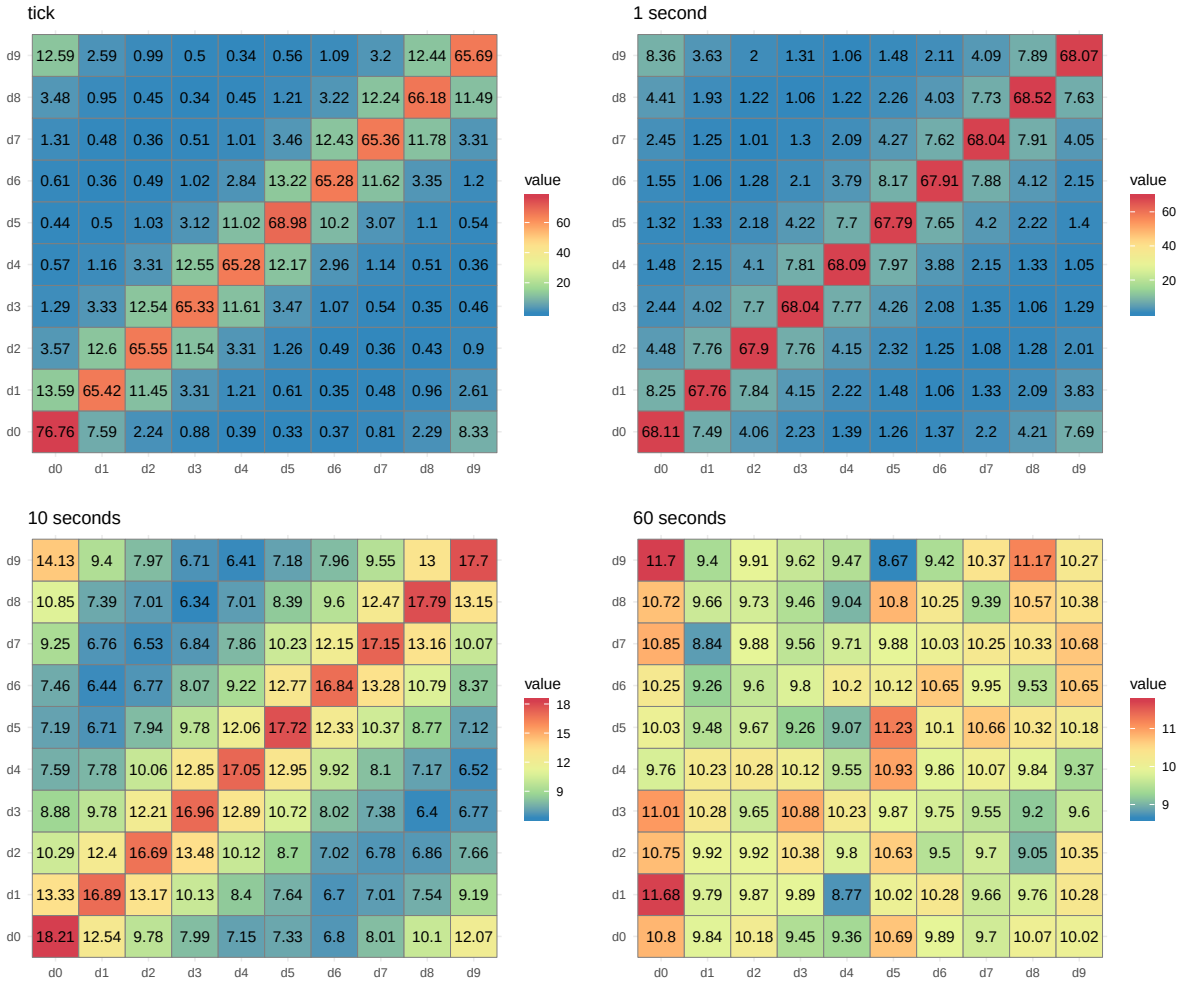


Figure 3 Conditional probabilities of decimals occurrence for AAPL

2.3 Conditional probabilities

In this section, a different data granularity is used to analyze the conditional probabilities of second decimals occurrence. The first heatmap of Figure 3 shows the conditional probabilities for AAPL tick data. The $a_{ij} : i, j = 0, \ldots, 9$, elements of the heatmap represent probabilities of observing j-th digit in a stock price conditional on observing i-th digit in a stock price of the previous transaction. Formally,

$$a_{ij} = P\left(x_t^{[2]} = j \,\middle|\, x_{t-1}^{[2]} = i\right), \qquad i, j = 0, \ldots, 9,$$

where $x_t^{[2]}$ denotes the d-th decimal digit of price x of t-th transaction. The second, third and fourth heatmap depicts the conditional probabilities for 1-second, 10-second and 1-minute AAPL data respectively in the same manner, however, $x_t^{[2]}$ denotes the d-th decimal digit of price x at time t.

The heatmap for the tick data shows that the conditional probabilities that digit 0 remains digit 0 and digit 5 remains digit 5 in the consequent transaction is distinctively higher (76.76% and 68.98% respectively) than the probabilities that other digits remain the same (ranging from 65.28% to 66.18%). Moreover, the probability of observing 0 after 1 is higher (13.58%) than observing 2 after 1 (11.45%). The same applies for digit 0 vs 8 after digit 9 (12.59% and 12.44% respectively).

When aggregating the tick data to 1-second data, the differences become less profound. However, it is still more likely that the 1 goes down to 0 than it goes up to 2 (8.25% vs. 7.84%) and that the 9 goes up to 0 than it goes down to 8 (8.36% vs. 7.89%) in the next transaction.

The heatmap for the 10-seconds data documents that the stronger price clustering in 0 digit shows up again. The probability that 0 remains 0 is the highest (18.21%). When aggregating the tick data to minute data, the conditional probabilities converge towards 10%. However, for 1 minute data we can still observe two distinctively higher probabilities than 10%: (i) probability of observing 0 after 1 (11.68%) and (ii) probability of observing 0 after 9 (11.70%).

3 Conclusion

This paper studied the price clustering phenomenon which refers to an excessive occurrence of transaction prices at certain digits. First, the state of the art on the price clustering phenomenon was summarized from its first observation at the NYSE in 1962 by M. F. M. Osborne, through the year 1994 when price clustering phenomenon has drawn much attention due to the study of NASDAQ stocks of Christie and Schultz, till the recent advances in knowledge of clustering phenomenon across different markets. Then, the paper took important steps towards proposing a new parametric model for high-frequency prices with discrete values since

- The data issue caused by wrong decimal places of precision was investigated. Based on an empirical study of four stocks traded on NYSE and four stocks traded on NASDAQ it seems that the prices should be rather rounded to the right decimal places instead of truncated.
- The price clustering phenomenon is persistent even when aggregating tick data to second and minute data.
- Both tick and 10-second data are suitable for parametric modeling which takes into account higher probabilities that 0 occurs in second decimal place after 0, 1 or 9 occurred as a second decimal of the previous transaction.

This paper contributes to the knowledge by investigating the (conditional) probabilities of price digits which evinces bumpy behavior due to the price clustering phenomenon. Our findings will be used for the future research of parametric modeling of high-frequency prices with discrete values.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/53/2019.

References

[1] Barndorff-Nielsen, O. E. & Hansen, P. R. & Lunde, A. & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Oxford University Press Oxford, UK*.

[2] Blasques, F. & Holý, V. & Tomanová, P. (2018). Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros. *arXiv preprint arXiv:1812.07318*.

[3] Christie, W.G. & Schultz, P.H. (1994). Why do NASDAQ market makers avoid odd-eighth quotes?. *The Journal of Finance*, 49(5), 1813–1840.

[4] Harris, L. (1991). Stock price clustering and discreteness. *The Review of Financial Studies*, 4(3), 389–415.

[5] Lien, D. & Hung, Pi-H. & Hung, I-C. (2019). Order price clustering, size clustering, and stock price movements: Evidence from the Taiwan Stock Exchange. *Journal of Empirical Finance*.

[6] Mishra, A. K. & Tripathy, T. (2018). Price and trade size clustering: Evidence from the national stock exchange of India. *The Quarterly Review of Economics and Finance*, 68, 63–72.

[7] Osborne, M. FM. (1962). Periodic structure in the Brownian motion of stock prices. *Operations Researche*, 10(3), 345–379.

Evaluation of trends in the development of expenditure on services in the Visegrad Group countries

Michał Bernard Pietrzak[1], Mateusz Jankiewicz[2]

Abstract. The article concentrates on the changes in the structure of service expenditure in the Visegrad Group countries. Therefore, in the research the Czech Republic, Poland, Slovakia and Hungary were under research, where the shares of expenditure on services were analysed in the years 1995-2012. After the accession of the Visegrad Group countries to the European Union, a systematic increase in socio-economic development was obtained. This has undoubtedly translated into an increase in expenditure on services in the final household spending. As a consequence, an increase of expenditure on services in the structure of consumer spending was observed in the period of the research. The main objective of the article is to determine the trends in the expenditure on services, where it was also attempted to set limit values for shares of this type of expenditure. The limit value was determined based on the estimation of parameters of Törnquist's function. The established limit value should be treated as a standard saturation level for expenditure on services for the Visegrad Group countries.

Keywords: Visegrad Group, consumption structure, service expenditure, Törnquist's function.

JEL Classification: C1, E2
AMS Classification: 62J02

1 Introduction

Services make an important element of any economy operating under the conditions of global globalization, and their volume and level of advancement indicates the level of development of the country. Research indicates that highly developed countries have the highest share of services in generating their GDPs. Services are also a significant part of consumer goods, which are demanded by households. Household spending on services is also largely linked to the socio-economic situation of the countries, and thus, the abilities of the housholds to generate additonal resources, which can be saved or spent on non-standard and non-basic consumption, see: [39], [41], [35]. Undoubtedly, the ongoing processes of digitization of economies and globalization affect the growth of consumption and changes in its structure, see: [24], [8], [9], [10]. Households in highly developed countries, with a higher level of disposable income, can afford to spend a larger proportion of their incomes on services compared to households from less developed countries. From the global perspective Chinese investments also generate a large investment expenditures as part of aggregate consumption in new EU member states [11]. Exchange rate fluctuations in CEE countries impact producer prices and finally consumer prices affecting household consumptioni pattern [15]. Therefore, it can be said that financial sector dynamics plays a significant role in the aggregate consumption movement in observed countries [33].

It should be emphasized that as part of the economic convergence process, the levels of domestic product per capita are becoming similar, especially in the countries forming convergence clubs (e.g., EU member states, including the Visegrad Group countries, see: [4], [20], [16], [10], [34]). This means that also the level of disposable income in countries that are subject to the convergence process is also becoming equated [29], [6]. The fact of households growing more affluent in less developed countries and the willingness to imitate the patterns taken from households in highly developed countries cause many social changes [17], where changes in the structure of consumer spending towards increasing the share of services expenditures are especially seen, see: [38], [32], [13], [14]. As a result, changes in the structure of consumer spending on services are a reflection of institutional changes, modernisation of the economies with bigger role of digital economy and international connections of individual countries, see: [28], [22], [23], [2], [3], [30], [21], [36], [37], [18].

[1] Nicolaus Copernicus University, Department of Econometrics and Statistics, Poland, ul. Gagarina 13a, 87-100 Toruń, mail: pietrzak@umk.pl
[2] Nicolaus Copernicus University, Department of Applied Informatics and Mathematics in Economics, Poland, ul. Gagarina 13a, 87-100 Toruń, mail: m.jankiewicz@umk.pl.

Although acknowledged as a positive fact, see [7] the phenomenon of increasing household expenditures on services do not itself indicate the increasing living standards or higher quality of life. It is worth mentioning, that some of the expenditures on services must be directed on the maintaining acceptable living environment for the household, see [42].

The purpose of the article is to examine changes in the structure of household services expenditures in the Czech Republic, Poland, Slovakia and Hungary in the years 1995-2012. The implementation of the objective made it possible to designate for each of the countries the maximum value of the share of services expenditure in final household consumption expenditures. A model based on the Törnquist function was used to determine the maximum values.

2 Methodolgy

The survey concerns the shares of services expenditures in final consumption expenditures in the period 1995-2012 and was carried out based on the data obtained from the European Statistical Centre database. The share of household services expenditures in the overall sum of expenditures in individual countries of the Visegrad Group (the variable Y) was analysed. Moreover, the study used the level of the gross disposable income per capita (the variable X). One of the research objectives of the analysis is to determine the level of the saturation of the shares of services expenditures in individual countries. The level of saturation is understood as the share of services expenditure that can be achieved by households in the future, providing the current trend in changes holds. To determine the level of saturation of the share of services in the final household consumption, the Törnquist model was applied, see [43]:

$$Y_t = \frac{\alpha_0 \cdot X_t}{X_t + \alpha_1} + \varepsilon_t, \tag{1}$$

where Y_t is dependent variable (the share of services expenditures in the final household consumption expenditures), X_t is independent variable (the level of disposable income per capita), α_0 is a parameter responsible for the saturation level, α_1 indicates the nature of a specific good, and ε_t is a random component.

Due to the non-linear form of the equation (1), the linearized form of the auxiliary model was used to estimate the parameters of the Törnquist model, see: [25]:

$$Z_{1t} = \beta_0 + \beta_1 \cdot Z_{2t} + \eta_t, \tag{2}$$

$$Z_{1t} = \frac{1}{Y_t}, Z_{2t} = \frac{1}{X_t}, \beta_0 = \frac{1}{\alpha_0}, \beta_1 = \frac{\alpha_1}{\alpha_0} \tag{3}$$

where: Z_{1t} is dependent variable, Z_{2t} is independent variable, β_0, β_1 are parameters of the model and η_t is a random component.

The auxiliary model parameters defined by the equation (2) are estimated using the classical method of least squares. For the share of services expenditures, a positive value of parameter β_0 and a positive value of parameter β are expected. The parameter α_0 indicating the level of saturation was determined based on equation, while the parameter α_1 indicating the nature of the good tested was determined in accordance with the equation

$$\alpha_0 = \frac{1}{\beta_0}, \alpha_1 = \beta_1 \cdot \alpha_0. \tag{4}$$

3 Empirical results

The survey was started with an analysis of the development of services expenditure and disposable income for the Visegrad Group countries for the 1995-2012 period[3]. Table 1 presents the value of the shares of services expenditure in final consumption expenditures and the level of household disposable income *per capita* in the Czech Republic, Poland, Hungary and Slovakia. Both in 1995 and in 2012, a higher and similar level of the shares of services expenditure was observed in the Czech Republic and Hungary and a lower and also similar level in Poland and Slovakia. In the case of the Czech Republic, Poland and Slovakia, the percentage change in

[3] Data are annual and disposable income is in current prices.

the shares of services expenditure exceeded 20%. In Hungary, the percentage change in the share of expenditure was 14.73%. The increase in the shares of services expenditure in final consumption expenditures is a consequence of effective modernization, growing competitiveness and the improvement of the basic macroeconomic indicators of the countries surveyed, see: [5], [8], [9], [19].

In turn, the level of disposable income *per capita* in 1995 was the lowest in Slovakia (1620.74 EUR), in Hungary and Poland it amounted to 2287.19 and 2020.47 respectively and reached the highest value in the Czech Republic (2478.33). It should be emphasized that in 2012 there were significant changes in the level of disposable income of individual countries. In the case of the Czech Republic and Slovakia, there was a much higher increase in the level of disposable income *per capita*, which exceeded EUR 8,000. However, Poland and Hungary, had a lower increase in the level of disposable income *per capita*, which exceeded slightly EUR 6,000. This fact confirms the diversity in the level of socio-economic development of individual economies.

Country	**Share of the services expenditures**			**Disposable income p.c.**		
	1995	**2012**	**Percentage change**	**1995**	**2012**	**Percentage change**
Czech Republic	0.372	0.449	20.80%	2478.33	8454.28	241.13%
Hungary	0.376	0.432	14.73%	2287.19	6011.02	162.81%
Poland	0.327	0.404	23.43%	2020.47	6252.61	209.46%
Slovakia	0.326	0.406	24.52%	1620.74	8018.95	394.77%

Table 1 Change of shares of the services expenditures and disposable income per capita in the Visegrad Group economies

In the next step of the study, the auxiliary model parameters were estimated using equation (2) and the structural parameters of the Törnquist I type were determined by the equation (1). The obtained results are presented in Table 2. For each country, the auxiliary model had a high degree of matching to empirical data (the coefficient R^2 values) and statistically significant structural parameters β_0 and β_1. This allowed to calculate the assessment of the parameters α_0 and α_1 of the Törnquist type I model. Positive evaluations of the α_1 parameter were obtained in all of the countries under examination, which confirms the perception of the nature of services as goods of higher order. On the other hand, the values of the assessment of the parameter α_0 show the level of saturation for the shares of services expenditure in final consumption expenditures. The highest level of saturation was recorded for the Czech Republic (50.3%). Then in the case of Poland, the saturation level was 48.2%. Hungary and Slovakia noted a comparable level of saturation amounting to 45.8% and 44.4%, respectively.

Czech Republic			**Hungary**		
Parameter	**Estimate**	**p-value**	**Parameter**	**Estimate**	**p-value**
β_0	1.987	~0.00	β_0	2.181	~0.00
β_1	1700.647	~0.00	β_1	1091.673	~0.00
α_0	0.503	-	α_0	0.458	-
α_1	856.210	-	α_1	500.440	-
R^2		0.980	R^2		0.833
Poland			**Slovakia**		
Parameter	**Estimate**	**p-value**	**Parameter**	**Estimate**	**p-value**
β_0	2.076	~0.00	β_0	2.251	~0.00
β_1	1934.482	~0.00	β_1	1199.258	~0.00
α_0	0.482	-	α_0	0.444	-
α_1	931.895	-	α_1	532.875	-
R^2		0.814	R^2		0.9053

Table 2 The results of estimation and verification consumption structure models for the Visegrad Group countries

The obtained results of the estimation of the Törnquist type I model parameters were also presented in Figure 1. The visual analysis of Figure 1 allows concluding that the theoretical values of the model and the actual values of the shares for services expenditure in in final consumption expenditures are high. In addition, for each of the countries the saturation levels for the shares of services expenditures were marked with a broken line. The Czech

Republic and Poland are the most distant from the level of saturation. In the case of these countries, the increase in the share of this type of expenditure is most likely as a result of institutional development and the development of the digital economy of both economies, see: [3], [26], [27].

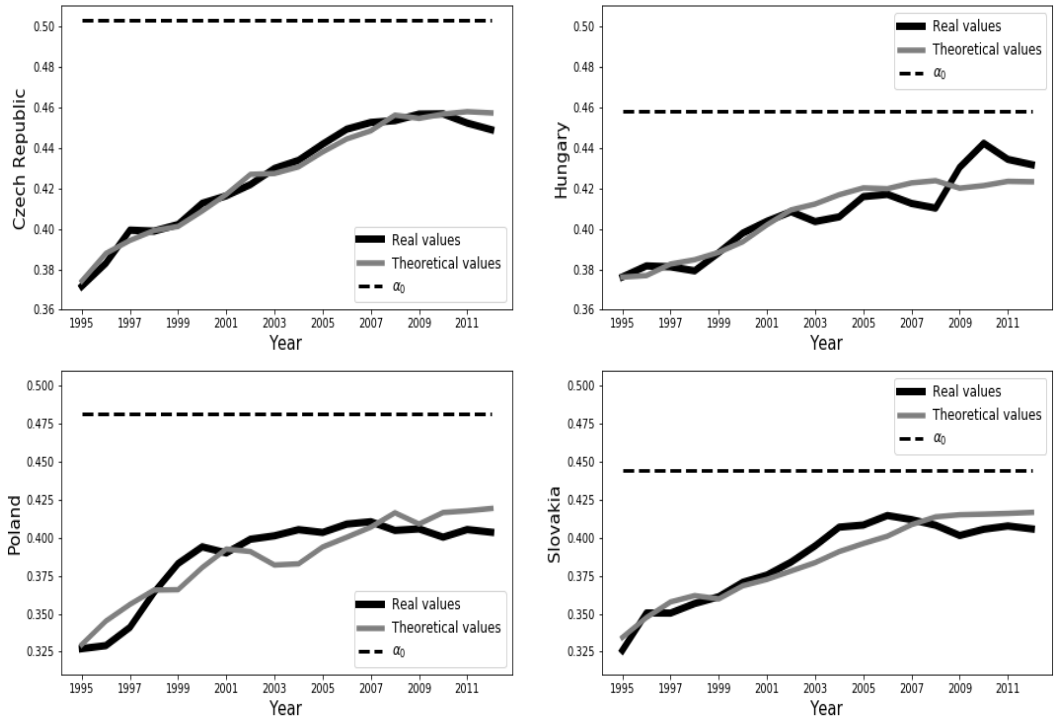


Figure 1 Comparing of the real and theoretical values of the share of the expenditures on services of households in the final consumption expenditures

4 Conclusions

The article analyses the changes in the shares of services expenditure in consumption structure in the Czech Republic, Poland, Slovakia and Hungary in the years 1995-2012. In the case of all of the Visegrad Group countries, there was an upward trend in the shares of services expenditures, which indicates a significant improvement in the economies of the countries surveyed. This means that these countries are becoming economically similar to the so-called old European Union countries, see: [1]. In the analysed period, a link was also established between the increase in disposable income *per capita* and the increase in the shares of services expenditures in final consumption expenditures. Undoubtedly, the changes in the household consumption pattern result from the increase in the level of socio-economic development of the Visegrad Group countries in 1995-2012, see: [30], [31]. This points to a significant link between macroeconomic indicators and household consumption of a specific good . However, it should be emphasized that the nature of the trend in the change in the shares of services expenditures is different for individual countries. This is a consequence of the different levels of advancement of the countries' economic development, including the level of institutional development, digital economy, *etc*, see: [26], [27]. By means of the Törnquist model, the maximum value of the share of services expenditures in consumption structure was also set for each country. The level of saturation ranges from 44% to 50% in all countries. The highest level of saturation was recorded in the Czech Republic. As a result of the analysis conducted, a slowdown in the upward trend in the shares of services expenditures in final consumption expenditures was noted for the Visegrad Group countries. However, it should be emphasized that the Czech Republic and Poland are much further away from the saturation level compared to Slovakia and Hungary. This means that Slovakia and Hungary should take steps to further increase the share of services expenditures to improve the functioning of the economy.

References

[1] Balcerzak, A. P. (2015). Europe 2020 Strategy and Structural Diversity Between Old and New Member States. Application of Zero-unitarization Method for Dynamic Analysis in the Years 2004-2013. *Economics & Sociology*, 8(2), 190-210.

[2] Balcerzak, A. P. (2016). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.
[3] Balcerzak, A. P. (2017). Digital Economy in Czech Republic, Slovakia and Hungary. Measurement with TOPSIS Based on Entropy Measure for Objective Weighting. In T. Loster & T. Pavelka (Eds.). *The 11th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016*. Prague: Libuse Macakova, Melandrium, 49-57.
[4] Balcerzak, A. P. & Rogalska, E. (2016). Non-Keynesian Effects of Fiscal Consolidations in Central Europe in the Years 2000-2013. In M. H. Bilgin, H. Danis, (Eds.) *Entrepreneurship, Business and Economics - Vol. 2. Proceedings of the 15th Eurasia Business and Economics Society*. Springer International Publishing, pp. 271-282,
[5] Cheba, K., & Szopik-Depczyńska, K. (2017). Multidimensional comparative analysis of the competitive capacity of the European Union countries and geographical regions. *Oeconomia Copernicana*, *8*(4), 487-504.
[6] Dykas, P., Mentel, G., & Misiak, T. (2018). The Neoclassical Model of Economic Growth and Its Ability to Account for Demographic Forecast Shocks. *Transformations in Business & Economics*, *17*, 2B (44B), 684 - 700.
[7] Fornell, C., Rust, R. T., & Dekimpe, M. G. (2010). The effect of customer satisfaction on consumer spending growth. *Journal of Marketing Research*, 47(1), 28-35.
[8] Gavurova, B., Vagasova, T., & Kovac, V. (2016). Competitiveness Assessment of Slovak Republic Regions. In J. Krajicek, J. Nesleha & K. Urbanovsky (Eds.). *European Financial System 2016. 13th International Scientific Conference of the European Financial Systems: Proceedings of the 13th International Scientific Conference*. Brno: Masaryk University (pp. 175-186).
[9] Gavurova, B., Soltes, M. & Kovac, V. (2017). Application of Cluster Analysis in Process of Competitiveness Modelling of Slovak Republic Regions. *Transformations in Business & Economics*, 16(3), 129-147.
[10] Głodowska, A. (2017). Business Environment and Economic Growth in the European Union Countries: What Can Be Explained for the Convergence?. *Entrepreneurial Business and Economics Review*, *5*(4), 189-204.
[11] Jacimovic,D., Dragutinović, R., Mitrović, Bjelić, P., Tianping, K., & Rajkovic, M. (2018). The role of Chinese investments in the bilateral exports of new E.U. member states and Western Balkan countries. *Economic Research - Ekonomska Istraživanja*, 31(1), 1185-1197.
[12] Jankiewicz, M. (2018a). Changes of the consumption structure in European countries considering its modernization process. In M. Papież & S. Śmiech (Eds.). *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings* (pp. 170-179.
[13] Jankiewicz, M. (2018b). The Influence of the Disposable Income on the Consumption Structure in European Countries – Spatio-Temporal Analysis. In *Proceedings of the 4th International Conference on European Integration 2018*, Ostrava (pp. 537-544).
[14] Jankiewicz, M. (2018c). Spatio-temporal analysis of the convergence of consumption expenditures in the European Union countries. In C. Brătianu, A. Zbuchea & A. Vițelar (eds.). *Challenging the status quo in management and economics : Strategica International Academic Conference, Sixth Edition*, Bucharest (pp. 142-154).
[15] Khan,K., Su, C.W., Tao, R., & Chu, C.C. (2018). Is there any relationship between producer price index and consumer price index in the Czech Republic? *Economic Research-Ekonomska Istraživanja*, 31(1), 1788-1806.
[16] Kisiała, W., & Suszyńska, K. (2017). Economic growth and disparities: an empirical analysis for the Central and Eastern European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(4), 613-631.
[17] Kliestik, T., Kovacova, M., Podhorska, I., & Kliestikova, J. (2018). Searching for key sources of goodwill creation as new global managerial challenge. *Polish Journal of Management Studies*, 17(1), 144-154.
[18] Ključnikov, A., Popesko, B., & Kloudová, J. (2019). Economics of the international ridesharing services - a trap for amateurs. *Entrepreneurship and Sustainability Issues*, 6(3), 1072-1081.
[19] Kruk, H., & Waśniewska, A. (2017). Application of the Perkal method for assessing competitiveness of the countries of Central and Eastern Europe. *Oeconomia Copernicana*, *8*(3), 337–352.
[20] Kuc, M. (2017). Is the regional divergence a price for the international convergence. The case of the Visegrad group. *Journal of Competitiveness*, 9(4), 50-65.
[21] Kurowska-Pysz, J., Szczepańska-Woszczyna, K., Štverková, H., & Kašík, J. (2018), The catalysts of cross-border cooperation development in euroregions. *Polish Journal of Management Studies*, 18 (1), 180-193.

[22] Łapińska, J (2013). The Business Sophistication and the International Competitiveness of the Polish Economy. In P. Jedlička (Ed.). *The International Conference Hradec Economic Days 2013, Economic Development and Management of Regions. Peer-Reviewed Conference Proceedings, Part III*. Hradec Králové: Gaudeamus (pp. 221-227).
[23] Łapinska, J. (2016). Determinant Factors Of Intra-Industry Trade: The Case Of Poland And Its European Union Trading Partners. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(2), 251-264.
[24] Markhaichuk, M., & Zhuckovskaya, I. (2019). The spread of the regional intellectual capital: the case of the Russian Federation. *Oeconomia Copernicana*, *10*(1), 89-111.
[25] Pawłowski Z. (1980). Ekonometria, PWN, Warszawa
[26] Pietrzak, M. B, & Ziemkiewicz, B. (2018a). Digital Economy in the Old European Union Member States. In T. Loster & T. Pavelka (Eds.). *The 11th International Days of Statistics and Economics. Conference Proceedings. September 6-8, 2018*. Prague: Libuse Macakova, Melandrium, (pp. 1431-1439).
[27] Pietrzak, M. B, & Ziemkiewicz, B. (2018b). Multiple criteria analysis of digital economy in the European Union Countries. In M. Reiff & P. Gezik (Eds.). *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XIX*. Trenčianske Teplice: Letra Edu, (pp. 283-290).
[28] Roman, A., Rusu, V. R., & Stoica, O. (2018). Institutional Environment and Economic Development in EU Countries: The Impact on New Business Formation Shocks. *Transformations in Business & Economics*, 17, 2B (44B), 579-603.
[29] Saman, C . (2018). The Predictive Power of Credit Fluctuations for Economic Activity in CEE-3 Countries. *Transformations in Business & Economics*, 17, 1(43), 178-191.
[30] Simionescu, M., Lazányi, K., Sopková, G., Dobeš, K., & Balcerzak, A. P. (2017). Determinants of Economic Growth in V4 Countries and Romania. *Journal of Competitiveness*, 9(1), 103-113.
[31] Simionescu, M., Balcerzak, A.P., Bilan, Y., & Kotásková, A. (2018). The impact of money on output in Czech Republic and Romania. *Journal of Business Economics and Management*, 19(1), 20-41.
[32] Shaw, S., Chovancová, M., & Bejtkovský, J. (2017). Consumer Behaviour and Warranty Claim: A Study on Czech Consumers. *Economics and Sociology*, 10(3), 90-101.
[33] Škare, M., Tomić, D., Porada-Rochoń, M. (2018). Exploring the Gibson law in CEE countries using a time series approach. *E & M Ekonomie a Management,* 21**(**4), 4-18.
[34] Stanišić, N., Makojević, N., & Tubić Ćurčić, T. (2018). The EU Enlargement and Income Convergence: Central and Eastern European Countries vs. Western Balkan Countries. *Entrepreneurial Business and Economics Review*, *6*(3), 29-41.
[35] Starostová, A., Zimmermann, P., & Stříteský, V. (2018). The Effect of Product Colour and Odour on Consumers' Perceived Preference and Intensity Ratings. *Journal of International Studies*, 11(4), 312-325.
[36] Stverkova, H., Pohludka, M., Kurowska-Pysz, J., & Szczepańska-Woszczyna, K. (2018). Cross-border enterprepreneurship in euroregion beskydy. *Polish Journal of Management Studies*, 18 (2), 324-337.
[37] Szczepaniak, I. (2018). Comparative advantages in Polish export to the European Union — food products vs selected groups of non-food products. *Oeconomia Copernicana*, *9*(2), 287-308.
[38] Szulc, E., & Jankiewicz, M. (2017). The labour market situation in medium-sized urban centres centres of the Kujawsko-Pomorskie voivodeship and the problem of unemployment in the province. In M. Papież & S. Śmiech (Eds.), The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings, 397-406
[39] Trębska, J. (2018). Polish households' savings in the financial intersectoral linkages. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *13*(2), 307-329.
[40] Wierzbicka, W. (2018). Information infrastructure as a pillar of the knowledge-based economy — an analysis of regional differentiation in Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *13*(1), 123-139.
[41] Verba, D., & Kudinova, A. (2019). Absolute value and diversity of household spending: analysis on International Comparison Program (ICP) 2011 data. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *14*(1), 99-127.
[42] Volodzkienė, L. (2018). The Diversity of Material Living Conditions (Housing Inequality): Lithuanian Case. *Intelektinė ekonomika*, 12(1), 59-72.
[43] Zalega T. (2011). Mikrouwarunkowania potrzeb i zachowań rynkowych gospodarstw domowych a nowa konsumpcja. *Studia i materiały*, 1-2/2011 (12-13), 79-106.

Analysis of unit trust funds and the creation of a descriptive model, with emphasis on the risk-factor and the return of investments

Oldřich Trenz[1], Oldřich Faldík[2], Sylvie Formánková[3], Jan Kolomazník[4], David Schubert[5], Daniela Kolomazníková[6]

Abstract The article deals with the data analysis of unit trust funds from the EU with ESG evaluation (environment, social responsibility and corporate governance), whilst attempting to categorize these funds from selected viewpoints. Emphasis is placed on identifying their risks and profitability, and this in connection with socially responsible investment (SRI), which includes sustainable investment (SI) which the ESG evaluation itself is a part of. ESG factors are part of the description of non-financial influences on long-term risk management and the expected return on investment. An analysis of the current approaches to evaluating investments shows that the key barrier in the transitions to SI is the fact that the ESG factors are not taken into account. A part of the article is also an analysis of the present state in the field of sustainable investments, the creation of one's own descriptive model, which takes into account ESG factors in the form of managing long-term risks and expected return on investments.

Keywords: socially responsible investment, sustainable investing, ESG, sustainability, descriptive model.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The beginnings of investing related to the development of society date back to the 18th century, when members of certain churches refused to invest into socially controversial companies; social values and religious communities were supported instead [4]. The very boom of socially responsible investment occurs in about the year 1970 together with raising environmental issues, such as processing of oil and waste, nuclear power plants, their waste and associated investments [3].

Corporate Social Responsibility (CSR) is a concept which was coined in the latter half of the 20th century in the context of growing expectations that corporations would assume their share of responsibility for the improvement of life quality in the sphere of their business [12]. The first pillar of socially responsible conduct was the environment (single-pillar conception), which was later supplemented with the social element (two-pillar conception) and today's variant also includes economy as the third pillar (3P concept – People, Planet, Profit) [14].

Similar principles as CSR began to develop in the 1970s in the area of investment. They are referred to as Socially Responsible Investing (SRI) [10]. This area includes various investment approaches and strategies and various financial products, which always have to be in harmony with the philosophy, according to which investors stress not only the expected profit in connection with the investment risk, but also investment into assets beneficial to the society, which respect the principles of sustainability [2], [6]. The two basic approaches to choosing the right portfolio include the principle of negative selection, which excludes tobacco, alcohol, gambling and arms industry (older approach) or positive selection, where the portfolio is made up of products of companies whose

[1] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.trenz@mendelu.cz.
[2] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.faldik@mendelu.cz.
[3] Mendel University in Brno/Faculty of Business and Economics/Department of Management, Zemědělská 1, 613 00 Brno, Czech Republic, sylvie.formankova@mendelu.cz.
David, P., Mendel University in Brno, Faculty of Business and Economics, Department of Accounting and Taxes, Zemědělská 1, 613 00 Brno, Czech Republic
[5] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, david.schubert@mendelu.cz.
[6] University of Applied Management, Cejl 40/107, 602 00 Brno, Czech Republic, daniela.kraydlova@globalmanagement.cz.

Articles of Association stipulate socially beneficial business and environmental development and sustainability [12]. Especially the last topic has been gaining in significance recently.

The socially responsible investment has been growing in connection with the growing offer of socially responsible unit trust funds. The expansion started in the US markets (in the 1970s) and spread to the United Kingdom (1980s); in the 1990s this concept became popular in Western Europe and after 2000 in Asia [16].

The main sectors in the offer of SRI funds are China (Asia), USA (North America) and Europe. Every fourth dollar invested into funds in the USA in 2017 complied with SRI principles. There was more than 38 % increase in the years 2016 to 2018, which resulted in the sum of USD 46.6 billion in 2018 [8]. Total assets in unit trust funds in Europe reached EUR 25.2 billion in 2018 [9]. The concept was also encouraged by the adoption of an EU directive applicable to companies employing more than 500 people. Such entities have to disclose descriptive non-financial information (environmental and social information), which means support for ESG descriptive information for funds investing into such companies [3].

Sustainable Investment (SI) is often discussed in connection with SRI. Such investing emphasizes factors associated with sustainable development, whereas environmental factors are given priority [5]. Investing which takes into account environmental protection and global warming prevention is referred to as green investing.

Investment portfolio which considers environmental aspects, social responsibility and corporate governance is referred to as ESG investment portfolio [1]. The environmental criterion includes aspects of environmental policy (utilization of local resources and treatment of waste), social aspects (occupational health and safety, working conditions etc.) and aspects associated with corporate governance (corporate conduct, equal opportunities, relationships within the company) [5]. Wherever the ESG factors are applied to financial products (investment funds), the underlying philosophy must be complied with.

In 2018, Eurosif presented a study evaluating current sustainable investment within the EU (the study involved 79 % of the relevant funds). The study is being presented as a "landmark" and as a manifesto in support of the Action Plan on Sustainable Finance supporting European policy changes in the financial sector [13]. The study confirmed that sustainable investment is very promising for individual investors who wish to contribute to the development of sectors concentrating on social aspects and environmental protection. Academic research affirms that there is a strong bond between ESG rating and financial performance in the long-term perspective [8].

It is appropriate to apply the sustainable investment aspects based on ESG rating in the investment decisions [17]. There are paid access databases where ESG criteria are calculated and taken into account for investment opportunities (investments into companies or funds). One of easily accessible databases of unit trust funds with ESG and environmental extensions is yoursri.com, which lists funds available in different countries. A common retail investor is financially limited to access the databases. This article aims at analyzing funds with ESG rating in the European Union and at designing a comprehensive evaluation model based on available parameters, which would be suitable for a new web project supporting retail investment.

2 Material and Methods

The next section will evaluate data concerning unit trust funds available in the European Union. Funds with ESG rating in individual EU countries will be selected. Emphasis will be put on the description of current situation in the area of affordable investment into unit trust funds with the ESG extension.

2.1 Input Data

The input data are data of capital funds with ESG factors available on the website of Socially Responsible Investment[7]. The advantage of this search engine is the possibility to apply the filter of the country in which the fund is traded. The sample of 11,513 SRI unit trust funds from 31 countries of the EU was the subject of evaluation. The data were collected during the year 2018.

ID	Ident. ISIN	Currency	Rat. ESG	Rat. ESG	SRRI	Total Assets	Performance: 1 Year	Performance: 3 Years	Performance: 5 Years	Since
X109	LU0229950067	Dollar	BB	3.9	4	61062400	-0.0183	-0.002	0.0210	6/ 2011
X110	LU0029876355	Dollar	BB	3.8	4	906802300	-0.1021	0.0223	0.0255	11/2013
X111	LU0170474422	Euro	BBB	4.7	3	121444000	-00502	-0.0099	0.0066	1/2011

[7] Socially Responsible Investment – https://yoursri.com/.

X112	LU0029871042	Dollar	BBB	4.5	5	2675656600	-0.0201	0.0076	0.0282	1/2011
…	…	…	…		…	…	…	…	…	…

Table 1 Selected descriptive data of ESG unit trust funds (subset: Czech Republic)

Table 1 shows a selection of descriptive data of unit trust funds, which will be subject to statistical examination. This is a limited range of descriptive indicators, which, for the purposes of this paper, have been reduced by short-term performance, by commodities or countries where the assets are allocated and also by additional risk indicators. The individual indicators are dimensionless numbers, the performance is a change of the fund value within the given period, expressed as percentage.

ESG Rating

ESG rating is a given number, without the methodology of its calculation (unpaid version of the database). The rating has a numeric value (0–10) as well as a mark (CCC, B, BB, BBB, A, AA, AAA). The higher the numeric value or the mark (CCC is the worst, AAA the best), the better quality of the fund in terms of ESG or sustainability [20].

SRRI

SRRI (Synthetic Risk and Reward Indicator) was applied for the assessment of the fund's risks based on the volatility of the fund. Volatility is determined on the basis of the fund's return for the past week; where the data are unavailable, the monthly returns of the fund are taken into account. The indicator is expressed by whole numbers from 1 to 7. A higher number means higher returns and higher risks, calculated according to formulas (1) and (2).

The returns relevant for the computation of volatility will be gathered from a sample period covering the last 5 years of the life of the fund and, in case of the distribution of income, will be measured considering the relevant earnings or the dividend payoffs. The volatility of the fund will be computed, and then rescaled to a yearly basis, using the following standard method:

$$Volatility = \sqrt{\frac{m}{T-1}\sum_{t=1}^{T}(r_{f,t} - \overline{r_f})^2}, \tag{1}$$

where the returns of the fund ($r_{f,t}$) are measured over T non-overlapping periods of the duration of 1/m years. This means $m = 52$ and $T = 260$ for weekly returns, and $m = 12$ and $T = 60$ for monthly returns; and where r_f is the arithmetic mean of the returns of the fund over the T periods:

$$\overline{r_f} = \frac{1}{T}\sum_{t=1}^{T} r_{f,t} \tag{2}$$

2.2 Evaluation of Data

Statistical Assessment of Input Data

Statistical software IBM SPSS[8] will be used for the assessment of the available data. Emphasis will be put on the identification of the relevant aspects suitable for further processing. The primary objective is to generalize the principles of this approach and apply them to a wider set of SRI unit trust funds worldwide.

Construction of Descriptive Model for Fund Evaluation

The next goal is the construction of a descriptive (classification) model for evaluating common funds with emphasis on identifying such funds that are suitable for creating investment portfolios according to select criteria. Although such evaluation models are available in the paid part of source databases of the yourSRI type, this however is not viable when creating independent instruments for the support of investment decisions (created web application) for small investors. A suitable version could be the creation of a composite index [7]; however, in the database without paid access there is not enough input for the calculation and the listed ESG marker is, in fact, already the required sustainability index. A much more interesting version is the employment of the ESG marker and other corresponding parameters describing common funds, e.g., the performance in the selected time-period (1, 3 and 5 years) and to include these outputs into the evaluation itself, into the construction of the descriptive model designated for the funds classification.

[8] IBM SPSS – https://www.ibm.com/cz-cs/analytics/spss-statistics-software

For building the evaluation model a self-organizing neural network (the learning without teacher type) seems suitable [11]. With the use of principles of creating clusters with the subsequent assessment, available common funds can be classified into required groups, based on the principle of the least distance from the representative of the classification group (there is a grouping of the inputs into sets based on similar characteristics of these inputs). This can be applied when creating a suitable portfolio, as part of the investment strategy (selection of funds of required characteristics; the risk-factor, the yield, the sustainability will be judged). Regarding the limited scope of this contribution, the self-organized neural networks and the used software will be referred to as part of the external resources [11], [19].

3 Results

Analysis of Data

For the primary EU funds analysis, we have divided the starting data sample based on its frequency corresponding to the given states and the re-counted percentage representation, see Tab 2.

Country	Freq.	%	Country	Freq	%	Country	Freq	%
Germany	985	8.6%	Nether.	649	5.6%	Greece	172	1.5%
Luxemb.	943	8.2%	Sweden	562	4.9%	Liechten.	159	1.4%
Switzer.	920	8.0%	Finland	521	4.5%	Czech R.	117	1.0%
France	871	7.6%	Belgium	418	3.6%	Hungary	100	0.9%
United K.	816	7.1%	Norway	415	3.6%	Poland	84	0.7%
Austria	800	6.9%	Ireland	413	3.6%	Slovakia	79	0.7%
Spain	762	6.6%	Denmark	402	3.5%	Estonia	67	0.6%
Italy	749	6.5%	Portugal	288	2.5%	Latvia	51	0.4%
						Others[9]	> 50	1.5%

Table 2 Distribution of the Sample in EU Countries

After the detailed analysis of the available data according to the sample, we have begun its reduction (for reasons of disambiguation), and this for reasons of incompleteness in some of the partial values. After the performed reduction to a complete subset, 957 SRI common EU funds were assessed.

On Fig. 1 you may see the situation of the distribution of risk-factors in the assessed funds. The majority is taken by the funds with an average risk-factor and thus also with lesser profitability. This confirms that the investors may be conservative regarding the SRI funds in Europe. Fig. 2 shows the 1, 3 and 5 years' performance. Here we can see that longer performance brings higher value. This can earn an SRI fund the reputation of a stable financial instrument suitable for long term investment with lower risk and lower reward. The situation shown on Fig. 3 displays the classified inputs – SRI funds – with the aid of a self-organizing neural network, into 3 groups (r1, r2 and r3). In this case, we are dealing with funds from the Czech Republic. The input data contained information on the ESG rating and performance for 1, 2 and 3 years.

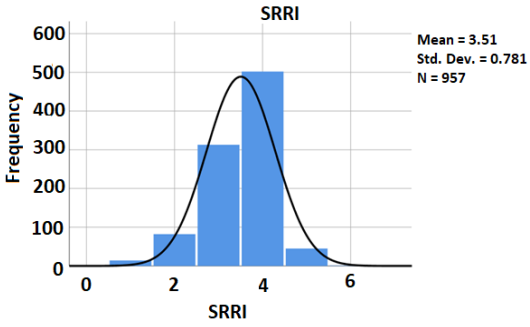


Figure 1 The risk-factor course – the SRRI (EU) indicator

[9] Others – Cyprus, Lithuania, Malta, Bulgaria, Romania, Slovenia, Croatia.

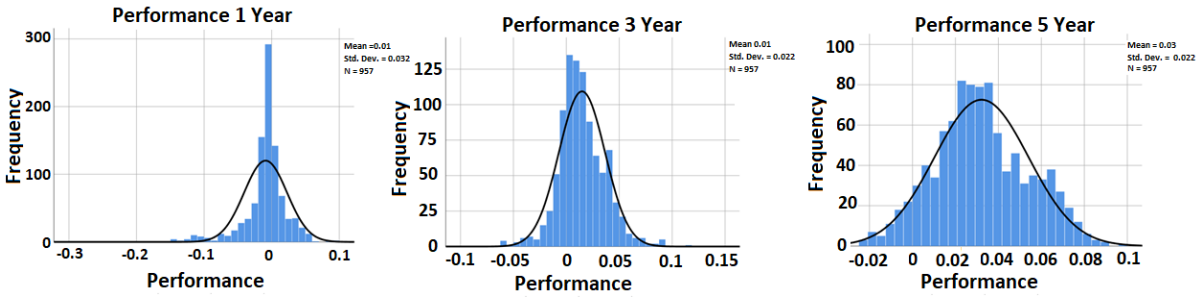


Figure 2 Evaluation of funds' performance after 1, 3 and 5 years (EU)

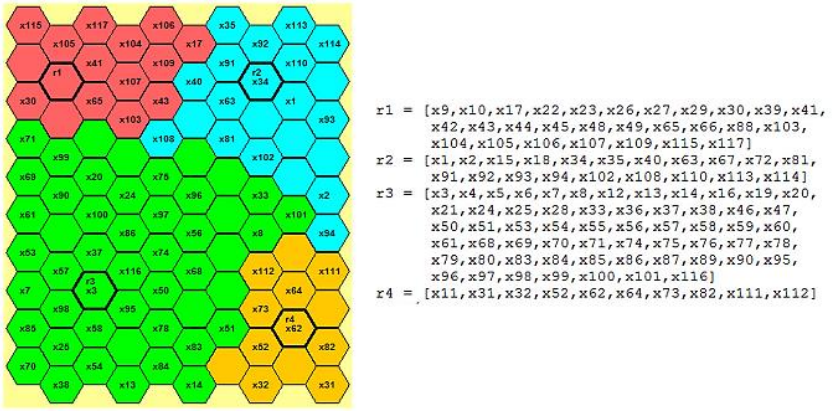


Figure 3 Funds classification using a neural network (Czech Republic)

For assessment (Fig. 3), software developed as part of the research organization (self-organizing neural network) was utilized. For more details, see for example [11].

4 Conclusions

The article describes the research which is part of a project focused on the evaluation of unit trust funds with elements of sustainable investment. Due to the limited scope of the contribution, we have substantially reduced the partial outcomes of the SRI fund evaluation in the EU and construction of a model for the evaluation of the funds prior to their inclusion into the investment portfolio.

The number of existing SRI funds and of institutions investing into such funds have increased and developed rapidly since the beginning of year 2000 [6]. Investors are likely to consider SRI due to economic, psychological and social reasons which are also attributed to the popular concept of sustainability. There has not been definite evidence of profitability of SRI; however, this is an important modern trend, supported by governments and bank institutions [15]. The research and evaluation of the relevant data made as part of this contribution suggest that SRI funds must be assessed in the long-term perspective of at least one year; the ideal investment appreciation comes after three to five years, as confirmed also by other authors [4].

Although the distribution of SRI funds among the EU countries is not even (Tab. 2), their performance is comparable and profitability similar, as mentioned in a number of studies. This is because the fund managers are centralized and offer the fund services in different countries. As results from the SRRI analysis (Fig. 1) show, the largest percentage is represented by medium-risk funds, which means that SRI funds are best suited for conservative investors. The described fact also follows from the evaluation of performance for 1, 3 and 5 years (Fig. 2). Another issue is the choice of the right approach of fund evaluation with regard to the intention to create the potential investor's portfolio. The funds may be selected on the basis of known facts (i.e. performance for a certain period – funds publish the data for one week, one month, 3 months, 6 months and for 1, 3 and 5 years; sustainability assessed by the ESG factor; and risk level in the form of SRRI).

The selection of a fund based on the calculation of a composite indicator and subsequent re-classification of funds according to its value has been found ineffective: such universal indicator cannot be formulated due to the various types of investors (conservative, those preferring higher risk and potential higher returns). Bearing in mind the above fact, we classified the funds with the help of a self-organizing neural network, while observing the

principle of similarity. Here, we can not only define the number of groups but also consider the underlying parameters of the classification. The input data for the classification of SRI funds in the Czech Republic (Fig. 3) were ESG rating and performance for 1, 3 and 5 years. SRRI may be used to check the correctness of identification of the groups. The findings of the research will be included into the web application with the support of decision-making, which aims at retail investors who contemplate sustainable investment into unit trust funds.

Acknowledgments

The authors are thankful to the Grant Agency of Czech Republic (GAČR), project: "Modelling and simulation of sustainable investment decision-making", CEP ID GA17-23448S.

References

[1] Ameer, R. (2016). Usefulness of ASSET4 ESG for Socially Responsible Investing in a Small Economy. *Journal of Investing*. 25(2), 125–135.
[2] Bengtsson, E. (2008). A History of Scandinavian Socially Responsible Investing. *Journal of Business Ethics*. 82(4), 969–983.
[3] Bonds and Climate Change. (2016). *The State of the Market Report 2016*. [cit. 2019-4-21] Available at: https://www.climatebonds.net/files/files/CBI%20State%20of%20the%20Market%202016%20A4.pdf.
[4] Bugg-Levine, A., Emerson, J. (2011). Impact investing: transforming how we make money while making a difference. San Francisco: Jossey-Bass. p 306. ISBN 978-0-470-90721-4.
[5] Combs, K. (2014). More than just a trend: the importance of impact investing. *Corporate Finance Review*. 18(6). 12–18.
[6] Cowton, C. J., Sandberg, J. (2012). Socially responsible investment. *In Ruth Chadwick (ed.)*. Encyclopedia of Applied Ethics. Academic Press.
[7] Dočekalová, M., Kocmanová, A. (2016) Composite indicator for measuring corporate sustainability. *Ecological Indicators*. 61(2016), 612–623.
[8] Duuren, E., Plantinga, A., Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *Journal of Business Ethic*. 138(3), 525–533.
[9] EFAMA's Asset Management report 2018 (2018). [cit. 2019-4-21] Available at: https://www.efama.org/Publications/Statistics/Asset%20Management%20Report/EFAMA_Asset%20Management%20Report%202018%20voor%20web.pdf.
[10] Forbes. (2017). *Socially-Responsible Investing: Earn Better Returns from Good Companies*. [cit. 2019-4-21] Available at: https://www.forbes.com/sites/moneyshow/2017/08/16/socially-responsible-investing-earn-better-returns-from-good-companies/#5542a29623d0.
[11] Konečný, V., Trenz, O., Svobodová, E. (2010). Classification of companies with assistance of self-learning neural networks. *Agricultural Economics*. 56(2), 51–58.
[12] Kunz, V. (2012). Corporate Social Responsibility. (Czech) Praha: Grada Publishing, 2012. p 201. ISBN 978-80-247-3983-0.
[13] Leaper, G. *Eurosif 2018 Study Finds Sustained SRI Growth*. (2018). [cit. 2019-4-21] Available at: https://nordsip.com/2018/11/27/eurosif-2018-study-finds-sustained-sri-growth/.
[14] Pícha, K. (2012). Social responsibolity of the company in the present trends and in contexts (Czech). Praha: Alfa nakladatelství. p 103. ISBN 978-80-87197-55-4.
[15] Rathner, S. (2012) The Performance of Socially Responsible Investment Funds: A Meta-Analysis, Working Papers in Economics 2012-2013, University of Salzburg.
[16] Salamon, M. (2014). *New frontiers of philanthropy: a guide to the new tools and actors reshaping global philanthropy and social investing*. Oxford: Oxford University Press. p 501. ISBN 978-019-935-7.
[17] Schramade, W. (2016). Integrating ESG into valuation models and investment decisions: the value-driver adjustment approach. *Journal of Sustainable Finance & Investment*. 6(2), 95–111.
[18] Song, Y., Wang, H., Zhu, M. (2018). Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR. *Financial Innovation*. Springer. 4(1).
[19] Trenz, O., Sepši, M., Konečný, V. (2014). Neural Network for Determining Risk Rate of Post-Heart Stroke Patients. *Acta of Mendel University of agriculture and forestry Brno*. 62(4), 769–776.
[20] USSIF - The Forum for Sustainable and Responsible Investment. (2018). *Performance & SRI*. [cit. 2019-4-21] Available at: https://www.ussif.org/performance.
[21] Waldman, A., Siegel, D., Javidian, M. (2004). *Transformational leadership and corporate social responsibility*. [cit. 2019-4-21] Available at: http://www.sristudies.org/Waldman+et+al+(2004).

Cooperative game with non-transferable utility and its benefits

Dana Figurová[1], Zuzana Čičková[2]

Abstract. The benefits of cooperative behavior can be observed in several ways. This contribution is dedicated into the distribution, especially the variant of Vehicle Routing Problem. In general, cooperative games can be divided into two groups according to its pay-off distribution: games with transferable and non-transferable pay-off. In this paper, we will deal with non-transferable pay-off in the case of mutual cooperation of the players, when these pay-offs are tied to the individual players and cannot be transferred to other players. It is considered the distribution system with customers who are assigned to an individual depots and the owners of individual depots (players) can cooperate in coalition to reach the savings of transport costs as well as the savings from mutual use of vehicles. The cooperation allows sharing customer's service and in that way it could bring the benefits from merging the transport requirements of individual players.

Keywords: cooperative game, Vehicle Routing Problem, non-transferable pay-off

JEL Classification: C70, C71, C60
AMS Classification: 91A12, 91A80, 90B06

1 Introduction

The Game Theory (GT) is a science that basically examines a wide range of decision situations for multiple participants. One way to explore these decisions is to accept the possibility of cooperation between the participants (players). The game is cooperative, if players can create the coalitions respecting the predetermined obligations and thus they can profit from a common approach. In non-cooperative games the players act independently and we do not consider such commitments.

We recognize the following types of conflict situations: (see in [11])

- Antagonist conflict: If two intelligent players participate in the conflict and choose their strategies to secure maximum winnings (pay-off), the win of one player is at the expense of the other participant (party games, military conflict).
- Non-antagonistic conflict: At least two participants (players) are involved in the conflict. The participants also maximize their pay-off. Winning one goes at the expense of others. If participants can enter into binding agreements on their decisions, we are talking about cooperative theory. If the winners are also redistributed, we talk about cooperative theory with a transferable utility. However, if the winnings cannot be redistributed, we are talking about a cooperative theory with a non-transferable utility.

Cooperative Games can be characterized as "Theory, which primarily deals with a coalition of players who coordinate their activities to achieve the further benefits" (see in [3]). In general, shipping costs represent a large part of the company's costs. One solution to reduce these costs is cooperation among the logistic companies. Zibaei et al. show [14], that the Cooperative Game Theory (CGT) can be adopted for modelling the cooperation among the companies. Cooperative Games are concerned with distribution of the cooperation benefits when the players cooperate. Most applications of the CGT are in scheduling, cost saving, negotiation and bargaining (see in [2], [10]).

In general, the cooperative games can be divided into two groups according to their distribution of pay-off. The first is games with non-transferable utility (pay-off). It is characteristic for games with non-transferable utility that the pay-offs are tied to the individual players and their benefits cannot be transferred to other teammates. An example of such a game is supporting a candidate in a contest that may have a limited number of players (eg, election of a president). Supporting a foreign candidate does not directly make a benefit to other subjects in the

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1,Bratislava, 852 35,Slovak Republic, dana.figurova@euba.sk
[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1,Bratislava, 852 35,Slovak Republic, zuzana.cickova@euba.sk

game (only one candidate may become a president), but may result in a better result without coalition support. (see in [5]). The Non-Transferable Utility (NTU) value is a solution concept for multi-person cooperative games in which utility is not "transferable" (games "without side payments").). Introduced by Shapley in [13], it generalizes the Shapley value [12] for Transferable Utility (TU) games, I.e., games with side payments, representable by a coalitional worth ("characteristic") function. Many economic contexts are more naturally modelled by NTU than by TU games; and indeed, the NTU value has been applied with some success to a variety of economic and economic-political models (see in [1]). Many authors have dealt with this issue. E.g. [4] characterized the unique subgame perfect equilibrium outcome and so justify a new value for cooperative games with non-transferable utility. Chun [7] investigated the implications of the axiom of coalitional concavity for non-transferable utility coalitional form games or Inoue [9] proved that every compactly generated non-transferable utility (NTU) game can be generated by a coalition production economy.

This paper is devoted to the comparison of the situation of transferable and non-transferable pay-off in the Cooperative Vehicle Routing Problem. Obviously, in the case of transferable utility can be higher savings from cooperation achieved (because the coalition acts “as a single subject”). In this case additional attention need to be paid to the subtraction of the surplus. In this case, additional attention needs to be paid to the subtraction of the surplus. In case of non-transferable prizes, this condition must be incorporated directly into the mathematical models. The proposed model includes a condition of individual stability, which means that a player cannot gain less than he would have achieved by a separate procedure.

2 Cooperative Vehicle Routing Problem with non-transferable utility

The presented model is an extension of a model introduced in [7]. We assume there are several centers (players), each of which has a set of customers. We assume the existence of multiple depots (players) associated in a coalition from which each of them has to serve a set of their own customers. We will formulate the mathematical model in the oriented and edge-weighted graph $\bar{G} = (N, \bar{H})$. Let: $N^{(1)} = \{d_1, d_2, \dots d\}$ is the set of nodes representing the central depots (centers). Depots offer the homogeneous goods in unlimited quantities. Let $N^{(i)}$ is a set of customers $of\ \ i-$ th depot, $i \epsilon N^{(1)}$ then $N^{(2)} = \bigcup_{i \epsilon N^{(1)}} N^{(i)}$ is a set of nodes representing the customers and $N = N^{(1)} \cup N^{(2)}$ is the set of all the nodes of the graph. We assume that there is only one type of vehicle in each depot with the vehicle capacity g and the number of vehicles in the depot is unlimited (or more than one round trip). Customer requirements $q_i \leq g$ for all $i \in N^{(2)}$. Compared to [7], we will assume that vehicles have to return back to its starting point (depot), after they came out from their depots. Let $\bar{H} \subset NxN$ represents the set of edges $e_{ij}, i, j \in N$ between all nodes *i* and *j*. Each edge e_{ij} being assigned to a real number called d_{ij}, also known as a price of the edge e_{ij}. Let this assignment be the shortest distance between nodes *i* and *j*. Each of the customers located at $i,\ i \in N^{(2)}$, require the import of a certain quantity of goods, generally denominated as $q_i, i \in N^{(2)}$. The goal is to determine the vehicle's routes, which will satisfy the requirements of all customers. Customer requirements will only be realized in the whole (if the vehicle serves the customer, its entire delivery requirement will be realized), with no vehicle capacity exceeded. The main goal is to minimize the total traveled distance. In the model we assume implicitly, that the size of each customer's requirement will not exceed the capacity of each depot´s vehicle. The aim of this paper is to model the situation of simultaneous customer service when it is not necessary to respect the initial allocation of customers to the individual depot, but another vehicle can be used to serve the customer in coalition.

In the model, we will uses only 2-index binary x_{ij} $(i, j \in N, i \neq j)$ and we accept the condition that vehicles have to return back to its starting point (depot), unlike the model presented in (see in [7]). We will also use in the model the non-negative variables $v_{ijh} \geq 0, i, j \in N, i \neq j, h \in N^{(1)}$ and even if they are not declared as binary, they will take the value $\{0,1\}$ and based on them it can be modelled, whether nodes (i, j) will be served by a vehicle from the *i*-th depot $i \in N^{(1)}$ or not. Let us recapitulate the model parameters and model variables more clearly

Sets and parameters

- $N^{(1)} = \{d_1, d_2, \dots d\}$ – set of centres (depots),
- $N^{(2)} = \{c_1, c_2, \dots c_m\}$ – set of customers (served nodes),
- $N = N^{(1)} \cup N^{(2)}$ – set of all nodes (customers and depots),
- $d_{ij}, i, j \in N, i \neq j$ – shortest distance moving from node *i* to node *j*,
- $q_i,\ i \in N^{(2)}$– demand of *i*-th customer,

- g – capacity of all vehicles,

Variables

- $x_{ij} \in \{0,1\},\ i,j \in N, i \neq j$ representing if the node *i* immediately precedes *j* in a route of some vehicle ($x_{ij} = 1$) or not ($x_{ij} = 0$).
- $u_i \geq 0, i \in N^{(2)}$ based on Miller - Tucker - Zemlin's formulation but representing vehicle load on its route to the *i*-th node (including).
- $v_{ijh} \geq 0, i,j \in N, i \neq j, h \in N^{(1)}$– variables representing the passing the vehicle of *h*-th depot throught the edge (*i*,*j*).

The mathematical model of Cooperative Vehicle Routing Problem can be written as:

$$cost = \min f\,(X,u) = \sum_{i \in N} \sum_{j \in N} d_{ij} x_{ij} \tag{1}$$

$$\sum_{i \in N} x_{ij} = 1,\ \ j \in N^{(2)},\ \ i \neq j \tag{2}$$

$$\sum_{j \in N} x_{ij} = 1, i \in N^{(2)},\ \ i \neq j \tag{3}$$

$$\sum_{j \in N^{(2)}} x_{ij} \leq 1,\ \ i \in N^{(1)} \tag{4}$$

$$\sum_{j \in N^{(2)}} x_{ij} = \sum_{j \in N^{(2)}} x_{ji},\ \ i \in N^{(1)} \tag{5}$$

$$u_i + q_j - g\left(1 - x_{ij}\right) \leq u_j,\ \ i \in N,\ \ j \in N^{(2)},\ \ i \neq j \tag{6}$$

$$q_i \leq u_i \leq g,\ \ i \in N^{(2)} \tag{7}$$

$$\sum_{h \in N^{(1)}} v_{ijh} = x_{ij},\ \ i,j \in N,\ \ i \neq j \tag{8}$$

$$v_{ijh} \leq 1 \tag{9}$$

$$v_{ijh} \geq x_{ij}, i \in N^{(1)},\ \ j \in N^{(2)} \tag{10}$$

$$v_{jih} \geq x_{ji},, i \in N^{(1)},\ \ j \in N^{(2)} \tag{11}$$

$$\sum_{i \in N} v_{ijh} = \sum_{l \in N} v_{jlh},\ j \in N^{(2)},\ \ h \in N^{(1)},\ \ i \neq j,\ \ l \neq j \tag{12}$$

$$\sum_{i \in N} x_{ijh} = \sum_{l \in N} v_{jlh},\ j \in N^{(2)},\ \ i,l \in N,\ \ i \neq j,\ \ l \neq j \tag{13}$$

The scalar *cost* (1) returns the total traveled distance. Constraint set (2) and (3) guarantee that each customer will be visited exactly once and exactly by one vehicle. Constraint (4) balance the use of vehicle of *i*-th depot, $i \in N^{(1)}$. Conditions (6) a (7) are the sub-tour elimination conditions and will also ensure that capacity of the vehicle is not exceeded. Conditions (8), (9), (10) and (11) modeled the service the edge with a from *i*-th depot $i \in N^{(1)}$. Continuity of the routes with vehicle of *i*-th depot is modeled by the constraints (12) and (13).

If conditions (8) to (13) are omitted from this model, the simplified model will allow to model the possibility of refilling the vehicle in a non-origin depot and this may lead to even greater savings (not just simultaneous customer service within the coalition but also simultaneous use of the depots).

Model (1) - (13) will allow to find the optimal solution of distribution for a given coalition. In this case, however, the condition of individual stability is not respected (i.e. the cost of an individual player may be higher than that which he would have in serving only his own customers). Let $f_k, k \in N^{(1)}$ represents the individual transport costs of each player. Their value can be quantified with using a capacitated vehicle routing problem (see in [7]) or using the model mentioned above where we will consider a single element set $N^{(1)}$. Thus, the following condition can be added to the model:

$$\sum_{i \in N} \sum_{j \in N} d_{ij} * v_{ijk} \leq f_k, k \in N^{(1)} \quad (14)$$

Thus, optimal solution of distribution within the coalition can be modeled for non-transferable utility (respecting the condition of individual stability). These mathematical models will be explained by illustrative example in the following section.

3 Results

Firstly, we are solving the cooperative model of Vehicle Routing Problem (1) – (13). The data were obtained from work paper (see at [6]), where we chose the symmetric distance matrix, which represents the distances of addresses between ten customers and two depots in Bratislava. We obtained the shortest distance matrix from origin data by Floyd algorithm.

Consider the net of 12 nodes. We will assume that there are 2 depots from which the vehicle can start its route. So, we consider the distribution problem with multiple depots, whereby we have 2 suppliers to serve the certain customers. Suppliers or players (owners of individual depots) are expressed as $N^1 = \{d_1, d_2\}$. Each player owns one depot with one type of vehicle. The same capacity of each vehicle is given by g = 200

Customers, who are strictly assigned to the individual depots (players), will be marked as*:* $\{c_1, c_2, c_3, c_4, c_5\}\, for\ d_1$, $\{c_6, c_7, c_8, c_9, c_{10}\}\, for\ d_2$. In the case of the creation the coalition $S \subseteq N^{(1)}$ we know that there are exactly 3 possible coalitions between the players including individual coalitions.

We solve the cooperative vehicle routing problem by using the model (1) - (13) and (1)-(14) for the created coalitions *S*: {1}, {2}, {1, 2} by GAMS software. To obtain the optimal solution, we used the solver Cplex 12.2.0.0 on the personal computer INTEL® Core ™ 2 CPU, E8500 @ 3.16 GB RAM for Windows 10. Our interest is to find optimal solutions by using the model of cooperative vehicle routing problem (1)-(13) ad (1)-(14) and compare their results. Model (1)-(13) will allow to find the optimal solution of distribution for a given coalition. However, in this case is the condition of individual stability not respected (i.e. the cost of individual player may be higher than that cost which the player has with serving only his own customers. In the model we accept the condition that vehicles have to return back to its starting point (depot), unlike the model presented in (see in [7]). Our main interest is to prove that there is a reduction in individual shipping costs through mutual cooperation between suppliers using the model (1)-(13) compared to the model (1)-(14) (with respect the condition of individual stability). In the Table 1, we can see the optimal routes of vehicles in each coalition and sum of the total costs.

Coaltions	**Optimal routes (1)-(13)**	**Optimal routes (1)-(14)**	**The sum of total costs of (1)-(13)**	**The sum of total costs of (1)-(14)**	**Costs of player 1 (1)-(13)**	**Costs of player 2 (1)-(13)**	**Costs of player 1 (1)-(14)**	**Costs of player 2 (1)-(14)**
S={1}	$\boldsymbol{d_1}$-c_3-c_5-c_4-c_2--c_1-$\boldsymbol{d_1}$	$\boldsymbol{d_1}$-c_3-c_5-c_4-c_2--c_1-$\boldsymbol{d_1}$	22,98	22,98	22,98		22,98	
S={2}	$\boldsymbol{d_2}$-c_8-c_{10}-c_9-c_7-c_6-$\boldsymbol{d_2}$	$\boldsymbol{d_2}$-c_8-c_{10}-c_9-c_7-c_6-$\boldsymbol{d_2}$	15,32	15,32		15,32		15,32
S={1,2}	$\boldsymbol{d_1}$-c_2-c_1-$\boldsymbol{d_1}$, $\boldsymbol{d_2}$-c_6-c_3-c_5--c_8-d_2-c_7--c_9-c_4-c_{10}-$\boldsymbol{d_2}$	$\boldsymbol{d_1}$-c_3-c_5-c_6-c_2- -c_1-$\boldsymbol{d_1}$, $\boldsymbol{d_2}$-c_8-d_2-c_{10}--c_4-c_7-c_4-$\boldsymbol{d_2}$	33,1	33,82	12,3	20,8	19,18	14,64

Table 1 Results of model (1)-(13) and (1)-(14)

We can summarize various results from Table 1. In the model (1)-(13) we accept the condition that vehicles have to return back to its starting point (depot). However, the condition of individual stability is not respected (i.e. the cost of an individual player may be higher than that which he would have in serving only his own customers). Model (1)-(14) which respects the condition of individual stability will allow to find the optimal solution of distribution for a given coalition. We can see that there is a reduction in individual shipping costs through mutual cooperation between suppliers using the model (1)-(13) and (1)-(14). With respect the condition of individual

stability, we can see that individual costs of individual players have changed their values. The total sum of cost in optimal routes of model (1)-(14) for a coalition $S = \{1, 2\}$ is 33,82. Individual costs of players in these model have different distribution compared to the model (1)-(13) caused by the respect of condition of individual stability. In model (1)-(14), we can also see the reduction when we compared the individual costs of individual players and individual costs of players in coalition.

4 Conclusion

The presented models are an extension of a model introduced in [7]. The goal is to determine the vehicle's routes, which will satisfy the requirements of all customers. In the model we assume implicitly, that the size of each customer's requirement will not exceed the capacity of each depot´s vehicle. The aim of this paper was to model the situation of simultaneous customer service when it is not necessary to respect the initial allocation of customers to the individual depot, but another vehicle can be used to serve the customer in coalition. In the model (1)-(13) is the condition of individual stability not respected (i.e. the cost of an individual player may be higher than that which he would have in serving only his own customers). For this reason, the situation in which the optimal solution of distribution within the coalition can be modeled as a model with non-transferable utility (respecting the condition of individual stability). By comparing our results, we have taken the decision that our model with the extra assumption obtain a different results. In model (1)-(14), we can also see the reduction when we compared the individual costs of individual players and individual costs of players in coalition. In conclusion, we can say that after the comparison our models with and without the extra assumption, the results of the optimal paths of each player's vehicles have changed. The results of the individual costs of individual players has improved after with the respect of the condition of individual stability (they will pay less than or at least as much as they would pay if they were in a one-member coalition).

Acknowledgements

This paper is supported by the project VEGA 1/0351/17 Application of selected models of Game Theory in economic problems of Slovakia

References

[1] Aumann, R. J. (1975). On the Non-transferable Utility Value: A comment on the Roth-Shafer Examples. *Econometrica, 43*, pp. 611-646.

[2] Barron, E. N. (2013). *Game Theory: An introduction.* New York: John Wiley & Sons Inc.

[3] Branzei, R., Dimtrov, D., & Tijs, S. (2008). *Models in cooperative game theory*. New York: Springer Science and Business Media.

[4] Clippel, G. (2007). The procedural value for cooperative games with non-transferable utility. *Mathematical Social Sciences, 53*, pp. 46-52.

[5] Demúth, A. (2013). *Teórie hier a problém rozhodovania*. Filozofická fakulta Trnavskej univerzity v Trnave.

[6] Figurová, D. (2017). Models and Algorithms for Courier Delivery Problem. University of Economics in Bratislava. Faculty of Economic Informatics. Department of Operations Research and Econometrics. Avaible online at: https://opac.crzp.sk/?fn=docviewChild00188CEE. [Accessed 21 March 2019].

[7] Figurová, D. & Čičková, Z. (2018) Cooperation in Distribution. In Quantitative Methods in Economics : Multiple Criteria Decision Making XIX. International Scientific Conference. *Quantitative Methods in Economics : Multiple Criteria Decision Making XIX : Proceedings of the International Scientific Conference : 23rd May - 25th May 2018, Trenčianske Teplice, Slovakia* (pp. 112-118). Bratislava : Letra Edu.

[8] Chun, Y. (2004). On weighted Kalai-Samet solutions for non-transferable utility coalitional form games. *Games and Economic Behavior, 47*, pp. 257-267.

[9] Inoue, T. (2013). Representation of non-tranferable utility games by coalition production economies. *Journal of Mathematical Economics, 49*, pp. 141-149.

[10] Jindal, P. & Newberry, P. (2018). To bargain or not to bargain: The role of fixed costs in price negotiations. *Journal of Marketing Research, 55*, 832-851.

[11] Peško, Š. (2004). *Teória hier*. Avaible at: http://www.km.fpv.ukf.sk/upload_docs/20170219_202314__0.pdf [Accessed 30 Apr. 2019].

[12] Shapley, L. S. (1953). A Value for n-Person Games. In *Contributions to the Theory of Games II*, ed. by H. W. Kuhn and A. W. Tucker (pp. 307-317).Princeton: Princeton University Press.

[13] Shapley, L. S. (1969). Utility Comparision and the Theory of Games. In *La Decisions, Aggreration et Dynamique des Ordres de Preference,* Paris: Editions du Centre National de la Recherche Scientifique, (pp. 251-263).
[14] Zibaei, S., Hafezalkotob. A., and Ghashami, S. Cooperative vehicle routing problem: an opportunity for cost saving. *Journal of Industrial Engineering International, 12*, 271-286.

Evaluation of a Crisis Situation Based on Incomplete and Unsound Data

Michal Škoda[1]

Abstract. This paper deals with evaluation of a crisis situation caused by natural disasters and/or accidents in the first moments after the situation occurs. The correct and timely evaluation of a crisis situations is essential, which is very easy to understand, but not so easy to achieve. Especially in situations where the decision maker has limited information and the consequences of a wrong decision can be disastrous.
The proposed approach was created to help to evaluate a crisis situation with high accuracy when using incomplete and unsound data. The entire approach is based on fuzzy set theory, which is widely used in representing uncertain knowledge. The paper also focuses on the choice of suitable scales and method of fuzzification and defuzzification that should be used in this kind of evaluation.
First, fuzzy linguistic scales and criteria are chosen. After this a crisis situation is evaluated based on the chosen criteria. Based on new approach of using α-level cut the completeness and soundness of data is taken into account. At the end, the defuzzification is done and the situation is ranked.

Keywords: Decision making, fuzzy number, linguistic scale, α-level, criteria, crisis.

JEL Classification: C44, D70, D80
AMS Classification: 90B50, 62C86

1 Introduction

Crisis situations caused by accidents and natural disasters are nothing rare in today's world. Among accidents and natural disasters we could include for instance floods, earthquakes, whirlwinds, severe wind gusts, hailstorms, droughts, fires, epidemics, power and goods breakdowns, gas explosions, aircraft crashes, or power grid breakdowns [22].

Papers usually focus on crisis preparedness and crisis management evaluation as in [4], [20]. But there is not much written recently about evaluation of the crisis itself. It is very important to take the right and effective decisions, which can sometimes be hard to achieve. Almost all emergency situations are closely connected with uncertainty, which can take many forms [3], [1], [5]. For the purposes of this article, the level of uncertainty will be sufficiently represented by the quality of information received. More specifically, by completeness and soundness of information as it is split by Motro [14]. Many authors use fuzzy theory to work with uncertainty [10]. The appropriateness of fuzzy set theory in emergency response situations has been confirmed by Altay & Green III [2].

The aim of this article is to present a general evaluation approach for crisis situations that is user-friendly, flexible, precise, working with vague expressions, and able to work with incomplete and unsound data. The proposed approach is based on Fuzzy Set Theory [21].

The structure of the paper is as follows: the second chapter describes the used tools and methods with the main focus on fuzzy set theory. The linguistic scales and fuzzification and defuzzification methods are picked. The whole approach is described in detail in the third chapter. The fourth chapter provides a case study based on real crisis situation. In the conclusion, the proposed approach and results are evaluated.

2 Necessary Background

In this section, the main tools and methods are reviewed as necessary background.

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague 6, mskoda@pef.czu.cz.

2.1 Fuzzy Number

Fuzzy numbers are special fuzzy sets in a set of real numbers $\mathbb{R} = (-\infty, +\infty)$. They usually have a special shape. The linear fuzzy numbers and especially trapezoidal fuzzy numbers are the usually used numbers [15]. Therefore, the trapezoidal fuzzy numbers will be used in this paper. More specifically if $x_1 \leq x_2 \leq x_3 \leq x_4$, then the trapezoidal fuzzy number $A(x) = (x_1, x_2, x_3, x_4)$ is defined by the membership function $A(x)$ [11]:

$$A(x) = \begin{cases} 0 & if \quad a < x_1 \\ \dfrac{a - x_1}{x_2 - x_1} & if \quad x_1 \leq a < x_2 \\ 1 & if \quad x_2 \leq a \leq x_3 \\ \dfrac{x_4 - a}{x_4 - x_3} & if \quad x_3 < a \leq x_4 \\ 0 & if \quad a > x_4 \end{cases} \tag{1}$$

The shape of the trapezoidal fuzzy number is crucial to the final outcome, since it is used to reflect the completeness and soundness of information. Assuming that the fuzzy linguistic scale is in ascending order and situations where the information is limited should be prioritized, then $(x_2 - x_1) < (x_4 - x_3)$ in case of the evaluation of a crisis situation and then $(x_2 - x_1) > (x_4 - x_3)$ in case of the evaluation of the completeness and soundness of the information. The greater the difference between $(x_2 - x_1)$ and $(x_4 - x_3)$ is, the bigger the impact will be on the final evaluation. Possible examples of such shapes of fuzzy numbers are presented in Figures 2,3 and 4.

2.2 Transformation of the fuzzy number

In this part the original idea of using α-level cut created by autor is presented.

Suppose we have a trapezoidal fuzzy number and the real number $\alpha \in \langle 0; 1 \rangle$ that will arrise from completeness and from soundness of information. The trapezoidal fuzzy number which kernel is equal to the α-cut of original fuzzy number will be called the transformed fuzzy number.

Definition:

The transformation of fuzzy number $A(x) = (x_1, x_2, x_3, x_4)$ is a fuzzy number $A_\alpha(x) = (x_1, x_{2\alpha}, x_{3\alpha}, x_4)$, where the element x_1 and x_4 remain the same and new values $x_{2\alpha}$ and $x_{3\alpha}$ are calculated using the following formulas:

$$x_{2\alpha} = x_1 + (x_2 - x_1) \times \alpha \tag{2}$$

$$x_{3\alpha} = x_4 - (x_4 - x_3) \times \alpha \tag{3}$$

Graphical representation of this transformation is presented in the Figure 1.

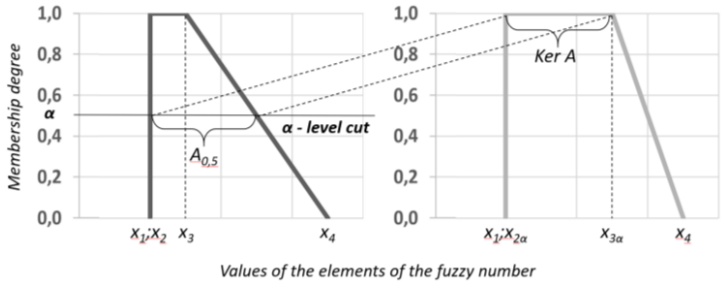


Figure 1 Transformation of trapezoidal fuzzy number

The transformed fuzzy number $A_\alpha(x) = (x_1, x_{2\alpha}, x_{3\alpha}, x_4)$ will be used as final parameter estimation. The lower the α, the bigger the impact on the final evaluation.

2.3 Fuzzy linguistic scales

Fuzzy set theory was developed to address the exact premise, that the key elements in human thinking are not numbers, but linguistic terms or fuzzy set labels [21]. Linguistic variables are descriptions employed by people. The relative values in a decision-making can be evaluated by linguistics terms such as 'good', 'fair', 'poor' and so on. These linguistics terms can be quantified and expressed as the fuzzy numbers [13]. In suggested approach,

the decision-maker should be asked to indicate their response category from linguistic scale where the endpoints are extremes, as for instance "Very Important" and "Unimportant". While creating the linguistic scale it is especially importance to consider number of scale categories and decide between balanced or unbalanced and odd or even scales [9].

For the purpose of this paper the balanced scale on universum $\langle 0, 1\rangle$ will be used. The number of favorable and unfavorable categories is equal. In some cases the special category "None" will be added. Also, we don't want the decision-maker to be neutral during evaluation of a crisis situation, so the neutral category could be excluded from the scale. The greater the number of scale categories is, the more accurate the resulting evaluation will be. On the other hand, most respondents cannot handle more than a few categories. As a compromise the five-part scales will be used (Figures 2,3 and 4). Such scales are easy to construct and administer and are easy for the respondent to understand [9].

The fuzzy scales used are defined at appropriate intervals $\langle u, v\rangle \subseteq \langle 0, 1\rangle$. The position of each fuzzy linguistic scale (interval $\langle u, v\rangle$) on universum is crucial to the resulting fuzzification and defuzzification. The further the fuzzy scale is from the origin, the bigger the resulting values are, and it is not necessary to use the weights of individual criteria.

2.4 Defuzzification

The goal of defuzzification is to interpret the fuzzy number as proper crisp value to be used by the decision-maker. (Sechilariu 2016). There are several methods of defuzzification. The center of gravity (COG) method stands out among them. This method returns the value of the center of area under the curve [8]. Additionally, COG method is simple and intuitively supportable, which raises the degree of confidence in the provided results [18].

Suppose $A(x)$ is a trapezoidal fuzzy number that is defined by four elements (x_1, x_2, x_3, x_4). Then for the COG $t_{A(x)}$ of fuzzy number $A(x)$, the following applies [12]:

$$t_{A(x)} = \frac{1}{3} \cdot \frac{x_4^2 + x_3^2 - x_2^2 - x_1^2 + x_4 x_3 - x_2 x_1}{x_4 + x_3 - x_2 - x_1} \tag{4}$$

3 Description of the evaluation approach

3.1 Criteria and linguistic scales

Initial step 1: Selection of Main Criteria

The first and essential step of this approach is the identification of criteria by which a crisis situation should be evaluated. For instance, a crisis situation can be evaluated based on the size of an affected area or based on a number of affected people [7].

Based on all possible types of crisis situations, four basic types of threats have been identified as main criteria by which crisis situations will be evaluated. The choice of these criteria corresponds with the Crisis Act [19] and are as follows: Life Threat, Health Threat, Threat to Property, Environmental Threat.

Initial step 2: Determination of Fuzzy Linguistic Scales

In this step it is necessary to define linguistic scale and related fuzzy numbers for evaluation of the crisis situation based on main criteria and for evaluation of completeness and soundness of the information received.

The fuzzy linguistic scale described in Table 1 was assigned to the Life Threat and Health Threat.

Linguistic term	Fuzzy number
None	(0.00, 0.00, 0.00, 0.00)
Very Low	(0.15, 0.15, 0.20, 0.40)
Low	(0.30, 0.30, 0.35, 0.60)
High	(0.45, 0.45, 0.50, 0.80)
Very High	(0.60, 0.60, 0.65, 1.00)

Table 1 Five-point scale

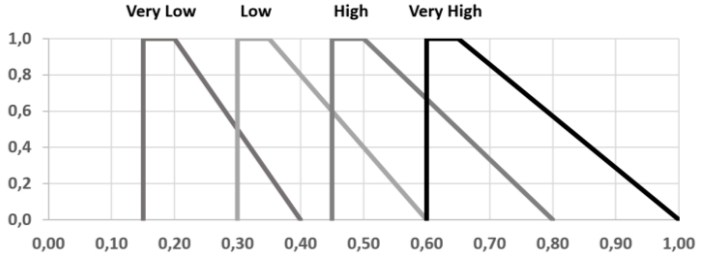


Figure 2 Five-point scale

Fuzzy linguistic scale described in Table 2 was assigned to Threat to Property and Environmental Threat.

Linguistic term	Fuzzy number
None	(0.00, 0.00, 0.00, 0.00)
Very Low	(0.00, 0.00, 0.05, 0.25)
Low	(0.15, 0.15, 0.20, 0.45)
High	(0.30, 0.30, 0.35, 0.65)
Very High	(0.45, 0.45, 0.50, 0,85)

Table 2 Five-point scale

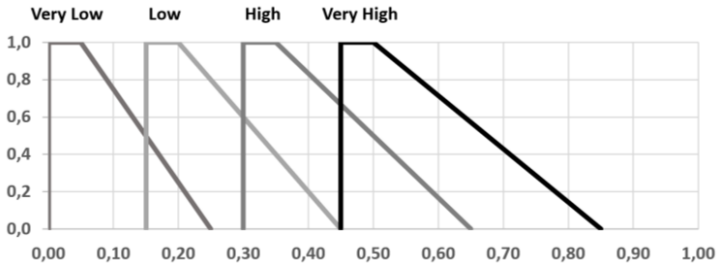


Figure 3 Five-point scale

The possible maximum value of COG that will be later use for the normalization is 0.784. The possible maximum is equal to the COG of the highest fuzzy number (0.60, 0.60, 0.65, 1.00), which correspond to the very high Life or Health Threat received after maximum possible transformation where the level of alpha is 0.192.

The evaluation of completeness and soundness of data will be done on the fuzzy linguistic scale described in Table 3.

Linguistic term	Fuzzy number
Very Poor	(0.00, 0.30, 0.35, 0.35)
Poor	(0.20, 0.50, 0.55, 0.55)
Acceptable	(0.40, 0.70, 0.75, 0.75)
Good	(0.60, 0.90, 0.95, 0.95)
Very Good	(1.00, 1.00, 1.00, 1.00)

Table 3 Five-point scale

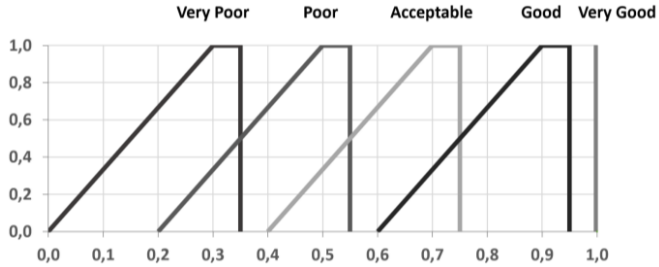


Figure 4 Five-point scale

3.2 Approach of evaluation of crisis situation

Step 1: Consideration of Soundness of Information

The soundness of information based on which the decision-maker evaluated the crisis situation should be evaluated on the pre-defined fuzzy linguistic scale.

Step 2: Consideration of Completeness and Soundness of Information

The completeness of information based on which the decision-maker will evaluate the crisis situation should be evaluated on the pre-defined fuzzy linguistic scale.

Step 3: Classification of a Specific Emergency Situation

If case a crisis situation appears, it should to be evaluated through the main criteria on the pre-defined fuzzy linguistic scale.

Step 4: Defuzzification and transformation of fuzzy numbers

First the fuzzy number evaluating the soundness of information should be defuzzified. The resulting number should be used as α-level to transform the fuzzy number resulting from evaluation of completeness of information.

Second should be defuzzification of the transformed fuzzy number resulting from evaluation of completeness of information. This value should be used as α-level to transform the fuzzy numbers resulting from evaluation of each main criteria based on which the crisis situation is evaluated.

Then the defuzzification of these transformed fuzzy numbers should be done. The resulting crisp numbers are normalized, so the results can be easily compared.

Step 5: Final Ranking

The final evaluation of the crisis situation is equal to the highest value resulting out of the previous defuzzification of the main criteria.

4 Case study

In this case study the designed approach for the evaluation is applied on major crisis situation that happened in Czech Republic a few years ago.

> **Railway accident in Studenka** - The international train from Prague crashed into the construction of a repaired road bridge that collapsed a few seconds before. Many passengers seem to be injured. Some of the passengers in the front part of the train may be dead. The partial information was provided by a shocked passenger from less affected part of the train. The situation will be evaluated as it could be in the first moments after the situations appeared.

Firstly, the possible evaluation of information received is presented in the Table 4.

Data Quality	Evaluation	Fuzzy Number	Transformed Fuzzy Number (α = 0.431)
Soundness	Poor	(0.20, 0.50, 0.55, 0.55)	-
Completeness	Good	(0.60, 0.90, 0.95, 0.95)	(0.60, 0.73, 0.95, 0.95)

Table 4 Evaluation of completeness and soundness of the information

The possible evaluation of the crisis situations is presented in the Table 5.

Fuzzy number assigned to poor soundness of data, based on Table 4 is (0.20, 0.50, 0.55, 0.55), which defuzzification by COG is 0.431.

Fuzzy number assigned to good completeness of data, based on Table 4 is (0.60, 0.90, 0.95, 0.95). The transformation of this fuzzy number based on $\alpha = 0.431$ is (0.60, 0.73, 0.95, 0.95). After defuzzification by COG, the resulting crisp value is 0.805. This crisp value will be used to transform the resulting fuzzy numbers from evaluation of the crisis situation.

The transformed fuzzy numbers of evaluation of each main criteria of the crisis situation and their normalization are in Table 5.

Threat	Evaluation	Fuzzy Number	Transformed Fuzzy Number (α = 0.805)	COG
Life Threat	High	(0.45, 0.45, 0.50, 0.80)	(0.45, 0.45, 0.56, 0.80)	0.575
Health Threat	Very High	(0.60, 0.60, 0.65, 1.00)	(0.60, 0.60, 0.72, 1.00)	0.742
Threat to Property	High	(0.30, 0.30, 0.35, 0,65)	(0.30, 0.30, 0.41, 0.65)	0.425
Environmental Threat	Very Low	(0.00, 0.00, 0.05, 0.25)	(0.00, 0.00, 0.09, 0.25)	0.091
			Maximum Value	0.742
			Normalized Max Value	0.946

Table 5 Evaluation of crisis situation

Based on Table 5, the resulting value of the crisis situation in the first moments after the situation occurred is 0.938. The value indicates the high severity of the situation. In the event that more crisis situations occur at once and the resources are limited, this procedure would help to prioritize.

5 Conclusions

During decision making it is very easy to forget that knowledge and assumptions are two different things as described Sherman and Harman [17]. Including the completeness and soundness of information into the calculations ensure that the difference between knowledge and assumptions will be taken into account.

It can be assumed that the suggested approach is generally usable for all types of crisis situations. The approach can be also considered as accurate, especially because the completeness and soundness of information is taken into account. It also could provide clear information about the seriousness of the situation to people who do not have enough time or knowledge to understand the situation. However, further research is needed to support these claims.

In the case study it was confirmed that the accuracy and relevance of the general overview of a crisis situation is improved using evaluation of the completeness and soundness of information. Also, the case study has shown that the whole approach can be applicable on a real crisis situation.

It should be noted that linear fuzzy numbers may not be sufficient in some cases. It would probably be necessary to use nonlinear fuzzy numbers for more complex and detailed evaluations.

Acknowledgements

This research is supported by the grant No. 2019A0018 “Využití kvantitativních přístupů v oblasti disaster managementu” of the Internal Grant Agency of the University of Life Sciences Prague.

References

[1] Afifi, Walid A., et al. “Uncertainty and Control in the Context of a Category-Five Tornado.” Research in Nursing & Health, vol. 37, no. 5, 2014, pp. 358–366., doi:10.1002/nur.21613.
[2] Altay, N., & Green III, W. G. (2006). OR/MS research in disaster operations management. European Journal of Operational Research, 175(1), 475-493.
[3] Benali, Akli, et al. “Fire Spread Predictions: Sweeping Uncertainty under the Rug.” Science of The Total Environment, vol. 592, 2017, pp. 187–196., doi:10.1016/j.scitotenv.2017.03.106.
[4] Boersma, F.K. ; Passenier, D.F. ; Mollee, J.S. ; van der Wal, C.N. / Crisis Management Evaluation: Formalisation & Analysis Of Communication During Fire Incident In Amsterdam Airport Train Tunnel. Proceedings of the 26th European Conference on Modelling and Simulation, ECMS, 2012. pp. 325-331
[5] Cheng, Yuan, and Xiaoping Zheng. “Effect of Uncertainty on Cooperative Behaviors during an Emergency Evacuation.” Communications in Nonlinear Science and Numerical Simulation, vol. 66, 2019, pp. 216–225., doi:10.1016/j.cnsns.2018.06.022.
[6] Cui, Lixin. “Applying Fuzzy Comprehensive Evaluation Method to Evaluate Quality in Crisis and Emergency Management”, Communications in Statistics - Theory and Methods, 41:21, 2012, pp. 3942-3959, doi: 10.1080/03610926.2012.691197
[7] Fang, Li, et al. “A Flood Emergency Response System Based on Flood Disaster Evaluation Models.” 2010 18th International Conference on Geoinformatics, 2010, doi:10.1109/geoinformatics.2010.5567665.
[8] Fuchs, Ewald, and Mohammad A. S. Masoum. Power Quality in Power Systems and Electrical Machines. Elsevier Science, 2015.
[9] Grover, Rajiv, and Marco Vriens. The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Sage Publications, 2006.
[10] Janál, Petr, and Miloš Starý. “Fuzzy Model Use for Prediction of the State of Emergency of River Basin in the Case of Flash Flood.” Journal of Hydrology and Hydromechanics, vol. 57, no. 3, 2009, doi:10.2478/v10098-009-0013-1.
[11] Klir, G. and Yuan, Bo. Fuzzy sets and fuzzy logic: theory and applications. Upper Saddle River, N.J.: Prentice Hall PTR., 1995. ISBN 978-0131011717.
[12] Leekwijck, Werner Van, and Etienne E. Kerre. “Defuzzification: Criteria and Classification.” Fuzzy Sets and Systems, vol. 108, no. 2, 1999, pp. 159–178., doi:10.1016/s0165-0114(97)00337-0.
[13] Lin, Hua-Yang, et al. “A Fuzzy-Based Decision-Making Procedure for Data Warehouse System Selection.” Expert Systems with Applications, vol. 32, no. 3, 2007, pp. 939–953., doi:10.1016/j.eswa.2006.01.031.
[14] Motro, Amihai, and Igor Rakov. “Estimating the Quality of Databases.” Flexible Query Answering Systems Lecture Notes in Computer Science, 1998, pp. 298–307., doi:10.1007/bfb0056011.
[15] Novák, Vilém. “Základy fuzzy modelování.” Prague: BEN - technická literatura, 2000. ISBN 80-7300-009-1.
[16] Sechilariu, Manuela, and Fabrice Locment. Urban DC Microgrid: Intelligent Control and Power Flow Optimization. Butterworth-Heinemann, 2016.
[17] Sherman, Brett, and Gilbert Harman. “Knowledge and Assumptions.” Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, vol. 156, no. 1, 2011, pp. 131–140. JSTOR, www.jstor.org/stable/41487725.
[18] Užga-Rebrovs, Oļegs, and Gaļina Kuļešova. “Comparative Analysis of Fuzzy Set Defuzzification Methods in the Context of Ecological Risk Assessment.” Information Technology and Management Science, vol. 20, no. 1, 2017, doi:10.1515/itms-2017-0004.
[19] Vaníček, Jiří. “Krizový zákon: komentář”. Prague: Wolters Kluwer, 2017. Komentáře (Wolters Kluwer ČR). ISBN 978-80-7552-787-5.
[20] Vichova, Kateřina, and Hromada, Martin. “The evaluation module of the crisis preparedness for the hospitals.” International Journal of Biology and Biomedical Engineering Volume 12, 2018, pp. 178-185
[21] Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338-353.
[22] Zuzák, Roman, and Fajfarová Martina. “Krizové řízení podniku”.Vlume 2., Prague, Grada, 2009. ISBN 978-80-247-3156-8.

Spatial Panel Data Analysis of Changes in Macroeconomic and Socio-demographic structures in Germany

Simona Macková[1]

Abstract. One of the strongest business drivers in Europe is Germany. Its economy is growing and being progressive in the past years even if it had to deal with the reunion in 1990. That time, former Western Germany was economically much stronger and had to support the weaker eastern part for a very long time. The gap between these two parts has definitely been narrowing but does it still exist or is the gap already bridged? Are the parts of Germany equal or did the inequality persist or was it transformed? The spatio-temporal nature of macroeconomic drivers and their dependencies should not be neglected. This contribution focuses on the spatio-temporal aspects and tries to observe and analyze the economic development of the country. Choropleths are an irreplaceable tool of spatial econometrics and are used to capture the economic situation and development in time visually. The paper uses analysis of data divided into NUTS-2 regions. As a measure of prosperity, gross domestic product, purchasing power standard, structure of industry and mean age of mother at childbirth are used. To evaluate assumptions about changes of economical structure, spatial econometric models suited for panel data are employed.

Keywords: spatial econometrics, spatial panel data, economic structure

JEL Classification: C21, C52, E66
AMS Classification: 91B72

1 Introduction

It has been nearly thirty years since Germany was reunited and the economically stronger western part started to support the weaker eastern. The economy is growing in the whole Europe and former Eastern Germany is growing as well. But can it ever approach or even overtake the former Western Germany? Let us take a look at the figure 1 which captures the situation in one decade between the years 2007 and 2016. The purchasing power standard per inhabitant (PPS) is an artificial currency unit used by Eurostat. One unit of this artificial currency could by the same amount of goods and services in each European country which reflects the well-being in the countries. PPS is measured in NUTS-2 regions separately. Just from the observation, one can be sure, that the PPS is increasing during the chosen time period. At the same time, lighter colour, that captures lower PPS, appears primarily in eastern and northern part of the country. The former boarder is still guessable.

The figure 2 shows the NUTS-2 regions division and over time changes of the mean age of women at childbirth. This measurement of fertility age reports the same trend as the PPS. The mean age of having babies is overall increasing but the gap between the former parts is still visible and does not seem narrowing much. And why the mean age of women at childbirth? Does it have anything common with economic prosperity? Already articles from the previous century as e.g. [1] and [8] pointed at the topic of fertility and economy. The age when women decide to have children is definitely influenced by the education standards. The higher education usually leads to better occupation and career which again postpones maternity. Both, [2] and [9] agree on this fact and also on the statement that higher educated population is one of the indices of the growing economy.

Based on these facts and assumptions, this contribution aims to analyze development of the economic prosperity of NUTS-2 regions in Germany regarding the historical separation. Since longer time period is observed and geographical structure is known, spatio-temporal methods would be the proper tool for this analysis.

This paper is structured as followed: firstly, the theory of spatial panel data models is briefly introduced and later the suggested models are applied on a chosen dataset. Results are interpreted and discussed with initial hypothesis.

[1] University of Economics, Prague, Department of Econometrics and Operational Research, W. Churchill Square 4, 130 67 Prague, Czech Republic, simona.mackova@vse.cz

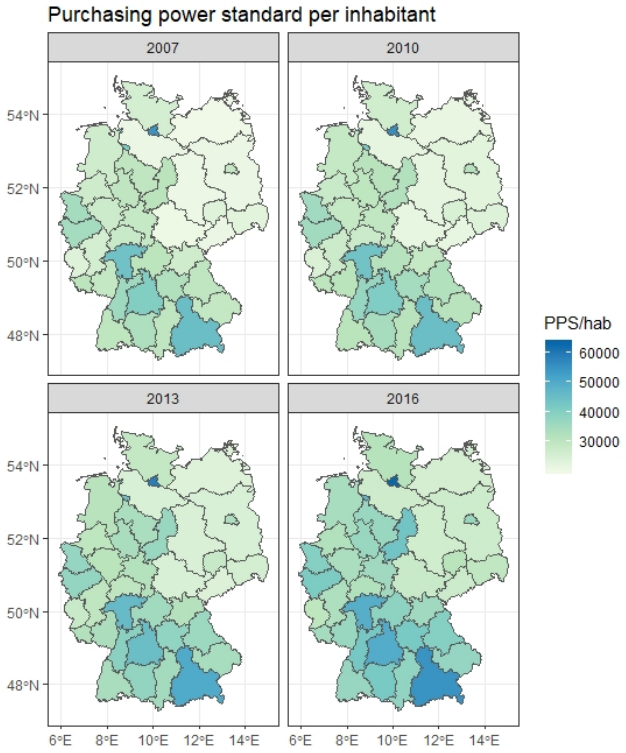


Figure 1: Purchasing power standard in Germany divided into NUTS-2 regions during the years 2007, 2010, 2013 and 2016

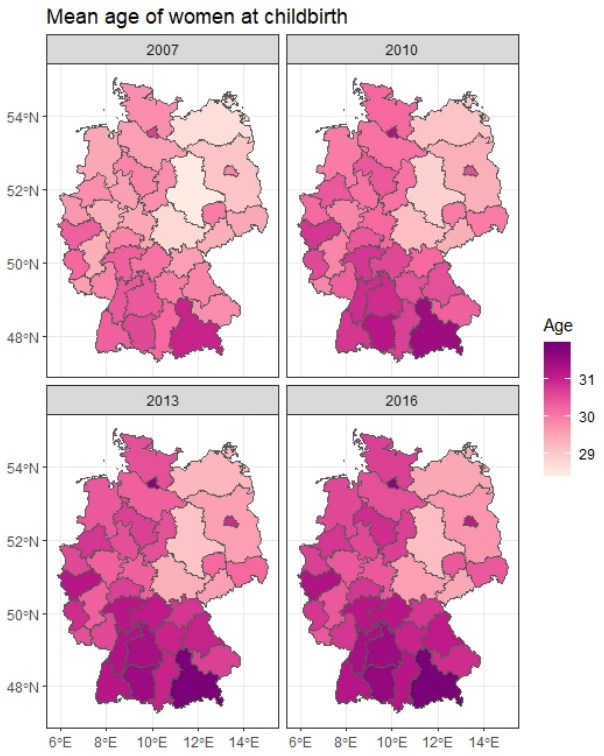


Figure 2: Mean age of women at childbirth in Germany divided into NUTS-2 regions during the years 2007, 2010, 2013 and 2016

2 Spatial Panel Data Models

Before starting to analyze the data, three crucial questions should be answered as Elhorst [3] suggests. Firstly, the kind of spatial interaction effect that should be considered. Secondly, one should decide whether to account spatial-specific, time-specific effects or both. If the time-specific effects are chosen, a decision whether to treat it as fixed or random effect has to be made.

The general nesting model for cross-section data considers all the possible spatial interactions. By dropping the spatial interactions, a wide range of spatial models is obtained. Such models used for N observations can be developed into a static spatio-temporal model suitable for panel data by adding t repetitions representing time running from 1 to T. The following spatial lag model is presented in [7].

$$\boldsymbol{Y}_t = \delta \boldsymbol{W} \boldsymbol{Y}_t + \alpha \boldsymbol{\iota}_N + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{u}_t. \tag{1}$$

In the previous equation 1, vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)^T$ represents the dependent variable and $\boldsymbol{X}$ stands for the matrix of regressors. $\boldsymbol{W}$ is the chosen spatial weight matrix representing spatial structure and $\boldsymbol{W}\boldsymbol{Y}_t$ is the endogenous interaction effect of the dependent variable. $\boldsymbol{u_t}$ stands for the vector of random part that is independently and identically distributed. The $N \times 1$ identity vector is represented by $\boldsymbol{\iota_N}$, where N in number of observed units. Further, $\boldsymbol{\beta}$ and scalars α as the intercept and δ as the spatial parameters will be estimated by a suitable tool.

The spatial weight matrix $\boldsymbol{W}$ is a deterministic structure that is not estimated in the model but it has to be constructed by the analyst at the beginning of the spatial analysis. Let us assume that the $\boldsymbol{W}$ is constant over the changing time period therefore the matrix has to reflects consistent socio-economic structure of the observed territory. Advanced methods considering changes of the $\boldsymbol{W}$ matrix are introduced for example by Lee and Yu [5].

Maximum likelihood estimator is commonly used tool for spatial panel data models estimation. For detailed description and derivation of spatio-temporal models and their estimators, see [6] or [7].

3 Empirical analysis

3.1 Data description

The whole dataset, that was used in this contribution, was provided by on-line database of European Statistic Office Eurostat (https://ec.europa.eu/eurostat/). As a dependent variable, purchasing power standard denoted as ***PPS*** was used. The gross domestic product was included among the explanatory variables as one of the most important macroeconomic measurements. While the production of industry is a significant part of each economy, variable ***IND*** was created as a ratio of persons employed in industry and total employed persons in the region. One can expect that industry the living standards increases in the region. The last regressor is denoted as ***FER*** and refers to mean age of women at childbirth. One would expect that increasing age of mothers could sign for well-being of the society. Reasons for including this variable were described in the introductory part.

To make this research reproducible, identification codes of used datasets are provided in the same order as above-mentioned variables: *tgs00026*, *nama_10r_2gdp*, *nama_10r_3empers*, *demo_r_find2*.

Also a dummy variable for NUTS-2 regions that used to belong to former Eastern Germany was created and denoted as ***EAST***. This binary variable should discover whether the former division of the country is still statistically significant for present economical prosperity.

All the variables are observed in the time period of ten years from 2007 until 2016. Since a balanced panel was required, the range of data with full information for longer period was limited.

3.2 Applied models

Since it is desirable to analyze impact of the dummy variable ***EAST*** that is constant in time, there is no point to consider fixed effect models. The basic theory about dummy variables and fixed effect models is well described e.g. in Wooldridge [10].

After the initial elaboration, it was decided to use a pooled spatial lag model for this contribution. The pooled specification is a type of model without individual effects. Similarly as in [4].

Let us denote a model for further elaboration.

$$\boldsymbol{PPS}_t = \delta \boldsymbol{W}\boldsymbol{PPS}_t + \alpha \boldsymbol{\iota}_N + \boldsymbol{GDP}_t\boldsymbol{\beta_1} + \boldsymbol{IND}_t\boldsymbol{\beta_2} + \boldsymbol{FER}_t\boldsymbol{\beta_3} + \boldsymbol{EAST}_t\boldsymbol{\beta_4} + \boldsymbol{u}_t. \quad (2)$$

for units $i = 1, 2, \ldots, 38$, where 38 is number of observed units - in this case regions in Germany divided according to NUTS-2 system, and for $t = 2007, 2008, \ldots, 2016$ representing annual observations. w_i stands for elements of spatial weight matrix constant over the time.

Several methods of constructing the spatial weight matrix $\boldsymbol{W}$ are available. In this contribution, the distance based method was chosen. Geographical centres were chosen as centroids. The Euclidean distances between centroids was measured and the threshold of maximal distance was set for 150 km which is the shortest distance that does not create any island (regions without any neighbours and connections). Further, coefficients were estimated using maximum likelihood method.

One of the initial hypotheses is the narrowing gap between former Western and Eastern Germany. Spatial lag models for cross-section data as described in [3] were applied on the same dataset and coefficients were estimated separately for years 2007 and 2016. These two estimations were compared and analyzed.

3.3 Results

The previous section introduces methodology used in empirical part. This section aims to present and discuss the obtained estimates and results.

The table 1 shows the coefficients estimated by the maximum likelihood method. All of them appear statistically significant and also meet the expectations. The gross domestic product, ratio of persons employed in industry and higher mean age of mothers at childbirth have positive impact on the purchasing power standard in the country. The dummy variable referring to eastern part of former Germany is also statistically significant but negative which states that these regions still fall behind economically.

As it is known from the basic theory of spatial analysis, estimated parameters as of models with spatial lag cannot be easily interpretable and direct and indirect effects have to be calculated. For derivation of the effects, see [3] or [6]. The effects, their z-scores and particular p-values are summarized in the table 2. The direct effects describe influence of increase of an explanatory variable on the dependent variable in the particular unit, indirect effects describe impact on the neighbouring units. Again, all the estimates are statistically significant and positive except the coefficient corresponding to the dummy variable ***EAST***. All the indirect effects are smaller than the direct which also makes sense: changes in a given unit have higher impact locally than on the neighbouring units. Results of the pooled spatial lag model are satisfying.

	Estimated coefficients
Intercept	-0.0003**
GDP	0.0085**
IND	4546.01**
FER	1445.20**
EAST	-820.53**
δ	0.2247**

** Statistically significant on 5% confidence level

Table 1: Estimated coefficients of the pooled spatial lag model

As suggested in the part describing methodology, spatial lag models for years 2007 and 2016 were calculated separately as cross-sectional models. Obtained direct and indirect effects of both models are summarized in table 3. Firstly, one observe that estimated coefficient δ is statistically significant only on 15% significance level in both cases. It could be discussed whether to use spatial models or not. Following, the main additional value of spatial models - indirect effect, are also statistically significant only on higher significance levels. The dummy variable ***EAST*** attracts the most of the attention. Its direct effects are statistically significant in both models which indicates that the gap between former Easter and Western Germany is still significant. On the other hand, the effect has decreased during in the observed ten-year period.

	Direct effects	Indirect effects	Total effects
GDP	0.0863	0.0238	0.1100
z-score	9.2462	4.4090	8.5277
p-value	0.0000	0.0092	0.0000
IND	4596.98	1266.13	5863.11
z-score	4.7399	3.3495	4.5969
p-value	0.0000	0.0734	0.0000
FER	1461.41	402.509	1863.92
z-score	12.6095	4.6051	10.8013
p-value	0.0000	0.0091	0.0000
EAST	-829.734	-228.53	-1058.26
z-score	-5.8367	-3.8175	-5.6639
p-value	0.0000	0.0106	0.0000

Table 2: Estimated direct and indirect effects of the pooled spatial lag model

	2007			**2016**		
	Direct effects	**Indirect effects**	**Total effects**	**Direct effects**	**Indirect effects**	**Total effects**
GDP	0.0997	-0.0190	0.0807	0.0626	-0.0124	0.0505
z-score	4.1071	-1.6538	3.5068	2.5855	-1.3970	2.3450
p-value	0.0000	0.0981	0.0004	0.0097	0.1624	0.0190
IND	2881.26	-549.10	2332.16	5522.64	-1070.44	4452.20
z-score	1.1304	-0.8314	1.1227	2.1760	-1.2914	2.0646
p-value	0.2583	0.4057	0.2616	0.0296	0.1966	0.0390
FER	1580.76	-301.26	1279.50	1445.8609	-280.25	1165.61
z-score	4.4414	-1.5854	4.3482	4.0511	-1.4995	3.9911
p-value	0.0000	0.1129	0.0000	0.0001	0.1338	0.0001
EAST	-968.370	184.55	-783.82	-839.17	162.65	-676.52
z-score	-2.3576	1.3767	-2.2441	-1.9427	1.2527	-1.8541
p-value	0.0184	0.1686	0.0248	0.0521	0.2103	0.0637
δ			-0.2283			-0.2238
p-value			0.1352			0.1083

Table 3: Estimated direct and indirect effects of the spatial lag model

4 Conclusion

This contribution described spatial panel data analysis in general, suggested suitable models and estimation tools for data structure including several observed units during chosen time period. Introduced methods were applied on a still topical theme of economic gap between former Western and Eastern Germany.

Purchasing power standard was chosen as a measurement of economical prosperity. Not only the gross domestic product and the ratio of population employed in industry appeared as significant drivers but also the age of women when they decide to have children. The main goal of this contribution was to discuss whether the parts of Germany are economically equivalent or whether the situation at the time of reunion is still persisting. Even though the gap seems narrowing, it is definitely still significantly present and it will obviously take very long time until it disappears.

For further research, suitable spatial dynamic models for panel data could be an interesting improvement of the analysis introduced in this paper. Also, following elaboration could take into account longer period from the history to observe trends and changes not only in the past decade.

Acknowledgements

This research project was supported by the grand No. IGA F4/60/2018, Faculty of Informatics and Statistics, University of Economics, Prague. Geo-data source: GISCO - Eurostat (European Commission), Administrative boundaries: ©EuroGeographics.

References

[1] Becker, G.S. (1992). Fertility and the Economy. *Journal of Population Economics*, 5(3), 185–201.

[2] Barro, R. (2013). Education and Economic Growth. *Annals of Economics and Finance*, 4(2), 301–328.

[3] Elhorst, J.P. (2014). *Spatial econometrics: from cross-section data to spatial panel*. Heidelberg: Springer.

[4] Formánek, T. & Macková, S. (2018). Spatio-Temporal Analysis of Macroeconomic Processes with Application to EU's Unemployment Dynamics. *Computational and Statistical Methods in Intelligent Systems*, 18, 120–131.

[5] Lee, L.F. & Yu, J. (2012). QLM Estimation of Spatial Dynamic Panel Data Models with Time Varying Spatial Weight Matrices. *Spatial Economic Analysus*, 7(1), 31–74.

[6] LeSage, J.& Pace, R.K. (2009). *Introduction to Spatial Econometrics*. U.S.A.: CRC Press.

[7] Millo, G. & Piras, G. (2012). spml: Spatial panel data models in R. *Journal of Statistical Software*, 47(1), 1–30.

[8] Mika, T. & Czaplicki, C. (2017). Fertility and Women's Old-Age Income in Germany. In: Kreyenfeld M., Konietzka D. (Eds.), *Childlessness in Europe: Contexts, Causes and Consequences* (pp. 331–349). Cham: Springer.

[9] Rindfuss, R.R., Bumpass, L. & John, C.S. (1980). Education and Fertility: Implication for the Roles Women Occupy. *American Social Review*, 45(3), 431–447.

[10] Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Trend analysis with use of binary representation

Michal Dominik Stasiak[1]

Abstract. One of the basic methods of technical analysis is the trend analysis. There are many indicators that allow for a current trend detection and parameter specification based on the candlestick chart representation. Because the use of candlestick representation is characterized by lower informative value, i.e. leads to a loss of information about the course trajectory, better modelling effects can be obtained by using a binary representation. The binarization algorithm transforms the course represented by tick data into a corresponding binary sequence. In the article are proposed algorithm allowing for a detection of current trend and specification of its parameters using binary representation. As a result of the algorithm performance we obtain, for each change in the binary representation, the type of the current trend and its parameters, resulting in a so-called binary-trend representation of a course trajectory. In the article we also include an exemplary statistical analysis of trends, performed based on historical tick data from a six-year period (2013-2019), for AUD/NZD currency pair. In order to execute the research a dedicated software was created in MQL4 and C++.

Keywords: foreign exchange market, high frequency econometric, technical analysis, trend analysis, modelling of currency exchange rates, currency market investment decision support.

JEL Classification: F31, G11, G14, C49
AMS Classification: 91G70, 62P20

1 Introduction

One of the most popular methods of investment decision support is the technical analysis. Many methods of technical analysis use trend analysis (e.g. visual techniques [2,3] or using dedicated indicators [8,12]). Methods used nowadays are burdened with two significant disadvantages which make it impossible to precisely describe the predictive potential of trend analysis. The first one is the lack of clear definitions of the start and end of a trend [12]– based on the same quotations two different investors can identify two different trends (e.g. visual analysis or subjective interpretations of indicators). Second disadvantage is using the candlestick chart representation, which leads to information that is unspecified and variable in time, that is to a loss of some crucial information about course changes within the candle [11,13]. In order to perform objective statistical research, allowing a description of relations between the trend and the probability distribution of future course trajectory changes, in this paper Author uses binary representation. Author also proposes an algorithm for an unequivocal distinction of the trend type and its parameters. The following paper also presents exemplary course trajectory changes analysis for AUD/NZD exchange rate, in a six-year research period, which proved the existence of dependencies between respective trend parameters and the future trajectory of the exchange rate quotations.

The paper is organised as follows. In the second chapter Author provides the description of the binary representation and justifies the possibilities of its application in technical analysis. In the third chapter an algorithm for identifying trend and its parameters is proposed for a binary representation of the course trajectory. The chapter also includes statistical analysis of researched currency pair. Chapter 4 summarises the most important from obtained results.

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, Niepodległości 10, 61-875 Poznań, Poland e-mail: michal.stasiak@ue.poznan.pl

2 Exchange rate representation

Exchange rates quotations change with very high frequency. Moreover, many changes of few pips length have the character of a noise [6,9]. Such a high frequency of changes and their noisiness require using proper means of course representation. Traditionally in the exchange rate analysis a very common candlestick chart representation is used [3,8,12]. In this kind of representation, trajectory changes for a given time period are represented by four variables: opening and closing course, and maximal and minimal price registered in the given time period. The candlestick representation is implemented in all broker platforms (e.g. Metatrader, JForex, etc.), and is widely used in research. Most of the technical analysis methods also uses the candlestick charts. An example can be visual analysis of waves developed by Elliot [2,5] or appointed indices such as RSI [14].

The lack of registering hundreds or even thousands of course trajectory changes of different amplitude within the „candle", which is described only by the four variables mentioned above, leads to a difficult to assess decrease of the informative value of data in the candlestick representation. The decrease is dependent on the frequency and amplitude of the course changes in the researched period. For example, in the night hours, significantly less changes of a smaller amplitude are registered in comparison to the high number of higher frequency changes registered during e.g. publication of important macroeconomic data. As a consequence, the unspecified and variable in time informative value of the candlestick representation of the exchange rate course trajectory results in a not very precise and inconclusive course analysis.

2.1 Binary representation

An alternative way to describe the course trajectory is the binary representation [11,13]. The idea of course binarization was first used to construct and analyse charts in point-symbolic method [4]. The basis for binary representation is the exchange rate discretization by a given change value, that is by so called "discretization unit" (DU). The binary representation of exchange rate quotations in a given time period is described by a respective binary string $\{\varepsilon_i\}_{i=1}^{n}$. The binarization algorithm assigns each ensuing element of ε_i value '0' if the course falls by one discretization unit, and value '1' if the course increases by one. In next steps, the algorithm calculates further binary values describing course changes in respect to the current quotation. Figure 1 presents an example of course discretization.

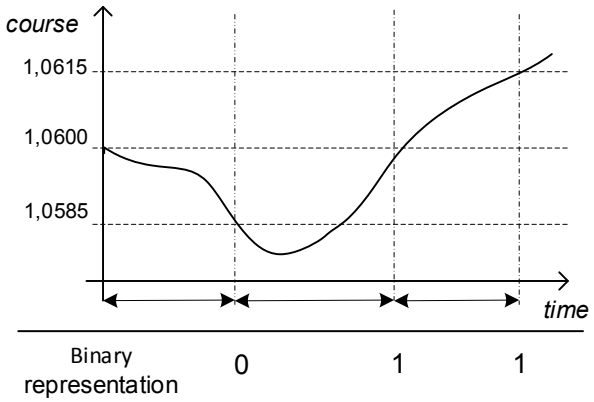


Figure 1 Exemplary tick data conversion in to binary data for the discretization unit of 15 pips

Let us now outline the analysis of changes character, if course trajectory changes of 15 pips range. In case of the candlestick representation, using even the smallest possible interval of 1 minute causes registering of thousands of changes of a very small (few pips) amplitude (e.g. at night), which bring nothing to the analysis. On the other hand, in the periods of intensive trading, one-minute candle can include even few changes of 15 pips. Therefore, using data expressed in a candlestick representation results in losing possibly important information about changes inside the candle on the one hand, or leads to the analysis of a high number of hardly important data on the other. As a consequence, the results of statistical analysis become unreliable. In case of using the binary representation, each significant change of the length described by the discretization unit will be registered, and smaller changes, having the character of a noise, will not be taken into account. If, based on so constructed representation, existence of relations between historical course trajectory changes and the future changes will be proven, then it would suggest that constructing proper prediction model in order to support investment decisions would be justified.

3 Trend in technical analysis

The term „trend" is one of the basic concepts used in economy. The existence of trends in quotations of different financial instruments is commonly accepted. Occurrence of trends can be easily explained by behavioural theories [10], differences in the time in which important information reaches each investor and so on. The general problem lies in something else – in the precise identification of the start and the end of the trend and in appointing its parameters. Investors are divided into two groups. First one adheres to a popular principle that "trend is your friend" and the second one plays to "break" the trend [8,12]. Most of the investors from both groups use subjective visual methods or different interpretations of some particular indicators in order to identify the trend. Without precisely defined rules for trend detection, a reliable statistical verification of the influence of the trend on the probability of future change direction is impossible. The problem begins with the very definition of the trend. The most general definition says that the trend "is a tendency to follow in a given direction in time" [8]. This definition is highly imprecise and does not allow for an unequivocal distinction of the beginning and end of a trend. Thus, the applicability of methods that are based on the trend analysis is dependent on the specification of exact rules of appointing the start and end of a given trend. Only those kind of methods would allow for a reliable statistical analysis of relations between the type of trend and its parameters and the direction of a future course change.

Using the candlestick chart representation of exchange rate quotations has a significant influence on the effectiveness of the trend analysis. From reasons provided in Chapter 2, the time-dependent loss of informative value of this kind of representation can lead, in periods of intense trading, to falsifying of results in cases of both visual analysis and statistical analysis of trend indicators such as RSI, Parabolic SAR [14] or MACD[1].

3.1 Identification of trend and its parameters in binary representation

We will introduce the following detailed definitions of trends, dedicated to the binary representation of a course trajectory.

Definition 1. An *increasing trend* is such a market state, in which for each ensuing *N* changes in the binary representation we observe a course increase of at least M pips (where M is a multiple of DU).

Definition 2. A *decreasing trend* is such a market state, in which for each ensuing N changes in the binary representation we observe a course decrease of at least *M* pips (where *M* is a multiple of DU).

Definition 3. A *horizontal trend* is such a market state, in which for each ensuing *N* changes in the binary representation the increase or decrease of the curse is not higher than *M* pips (where *M* is a multiple of DU).

Each trend can be assigned its type and two basic parameters: the duration and scope. In the binary representation, information about the change duration in the time units is omitted. Because of this, the duration of trend (*D*) in the binary representation is defined as follows:

Definition 4. The *duration of a trend* is expressed in the number of changes registered between the start and the end of a given trend.

In the other hand, the scope (*S*) in the binary representation is defined as:

Definition 5. The *scope of a trend* is expressed by the maximal increase (in case of the increasing trend) and the maximal decrease (in case of the decreasing trend) of the course, registered in relation to the beginning of the trend.

For the horizontal trend and for obvious reasons, we assume a constant scope which equals zero.

3.2 Trend-binary representation

Based on the detailed definitions presented in the subsection 3.1. we now propose a trend-binary representation, in which each course change by one discretization unit is assigned the current type, duration (in number of changes) and the scope of the trend. Figure 2 shows the block diagram for the algorithm used to construct the presented trend-binary representation.

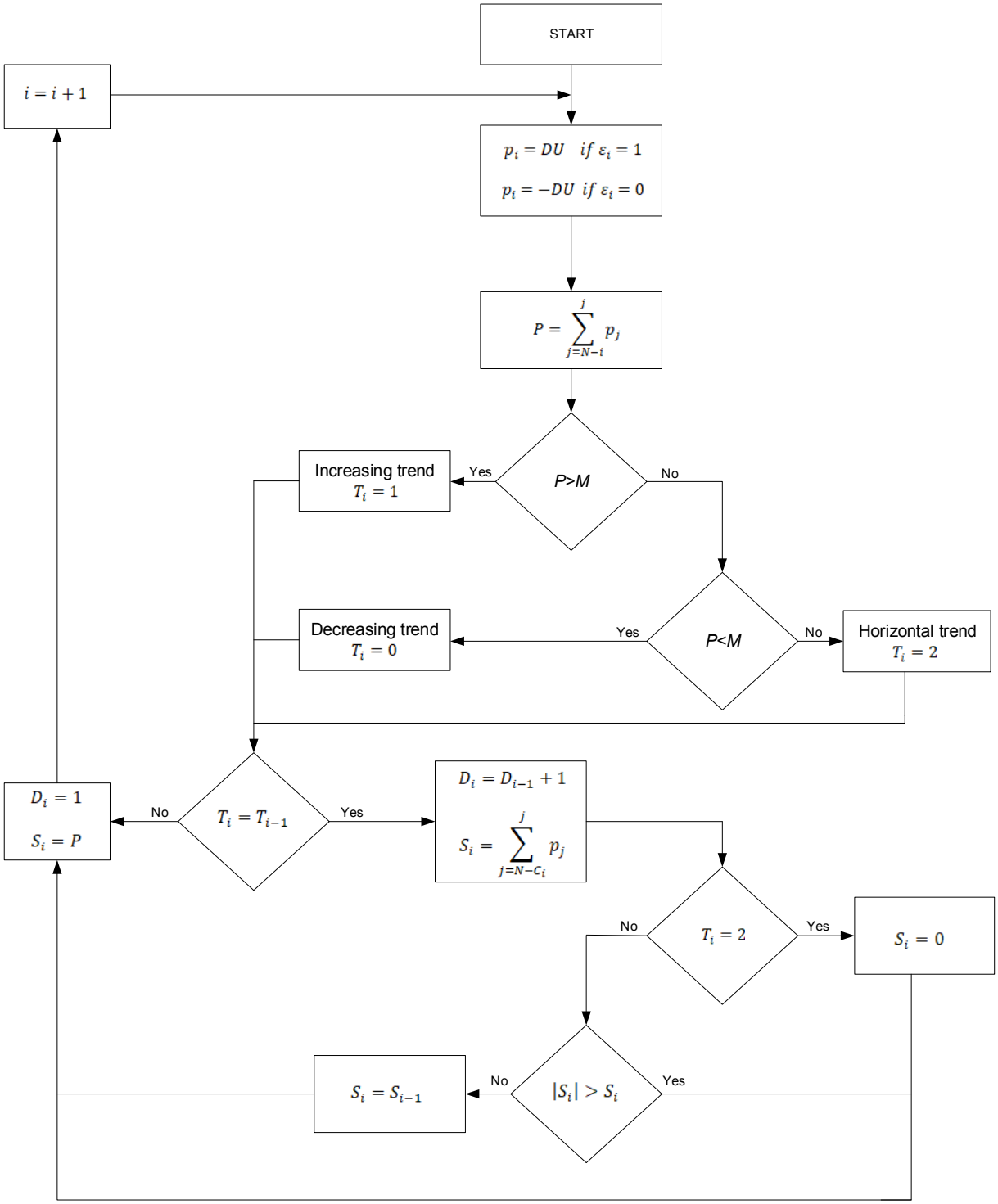


Figure 2 Block diagram for the algorithm used in constructing a trend-binary representation

The trend-binary representation allows for a statistical analysis, required for an objective verification of dependencies between the current type and parameters of the trend, and the direction of future course trajectory changes. The properties of the trend-binary representation depend on the assumed discretization unit (DU) and on the trend identification parameters (*N* and *M*).

3.3 Trend parameters analysis

Let us now consider an exemplary trend analysis for exchange rate AUD/NZD in the time period of six years (01.01.2013 – 31.12.2018), for an assumed discretization unit of 12 pips and parameters *N*=10 and *M*=24 (which mean that the trend will be detected if the increase in each 10 changes is equal to at least 24 pips).

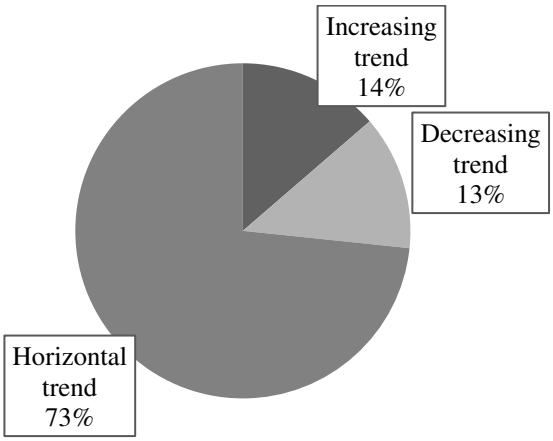


Figure 3 Percentages of registered trends of each type

Figure 3 presents the percentages of registered trends of each type. One can see that the percentage of the horizontal trend is the highest. This can stand as a proof that increasing or decreasing trends occur as an effect of publishing important information, and after the information is absorbed by the market, the course stabilizes on a particular level, which is characterized by small fluctuations in both directions, that is, in fact a horizontal trend. Research performed for other discretization units and trend identification parameters also give the same results.

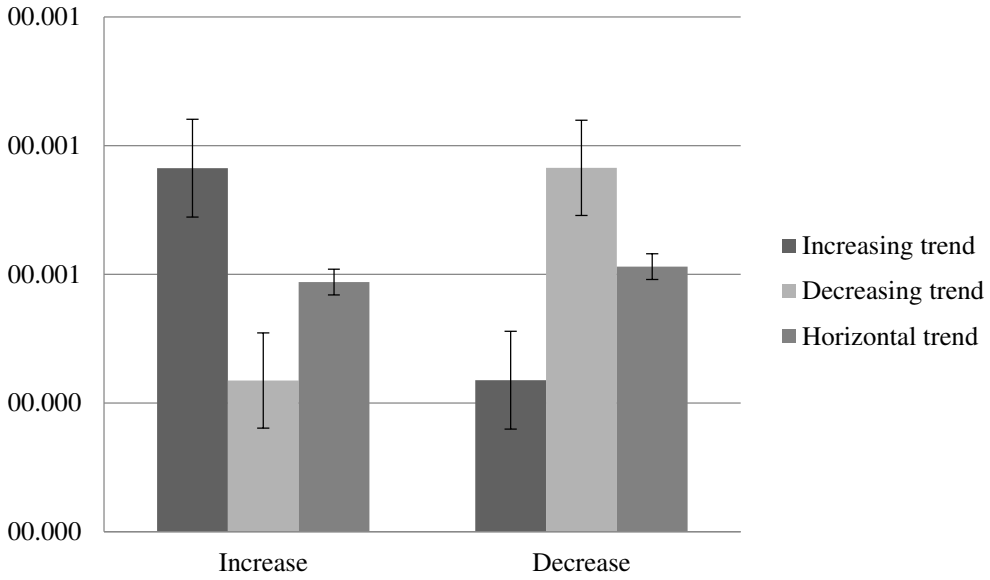


Figure 4 Probability of the future change direction depending on the trend type

Finally, let us consider the influence of the trend type on the probability of the future change direction. Figure 4 shows the probability distribution depending on the trend type. Results indicate a small but still statistically significant influence of the trend type on the probability distribution of a future change. The influence is more visible if we add the trend identification parameters to the equation. For example, the probability of an increase in case of an increasing trend longer than 10 changes equals 0.5706 +/- 0.021. On the other hand, if the increasing trend is longer than 7 changes the probability equals 0.5601 +/- 0.023. Presented results show that the analysis of a properly defined trend has a great prognostic potential. The trend type influences the probability distribution of the future course trajectory change in case of the increasing and decreasing trend (the horizontal trend provides no prognostic value). Introduction of trend identification parameters to the course trajectory analysis leads to more precise results. In consequence, the investor using the trend analysis can have even 7%-high advantage over the general market.

4 Summary

The trend analysis is one of the basic methods of technical analysis, used commonly as an investment decision support tool. Methods applied nowadays have two serious disadvantages which prevent a reliable statistical verification: lack of unequivocal and precise rules for identifying the start and end of a trend, and using candlestick representation, which can lead to a loss of informative value of data and makes it less useful in investment decision support.

The paper presents an algorithm for identifying a trend and its parameters in a binary representation. All algorithms presented in the article allow for construction of a trend-binary representation, which can be used to perform statistical analyses and to describe the influence of the trend and its parameters on the probability distribution of future course change direction. Additionally, the article presents exemplary results of research done on six-year quotations of AUD/NZD exchange rate. Mentioned research proved existence of an influence of the trend and its parameters on the probability of the future course change direction. Presented results justify the possibility of using the trend analysis to construct automatic HFT systems, characterised by a positive return rate.

References

[1] Appel G.: *The Moving Average Convergence Divergence Method*, Great Neck, NY: Signalert, 1979.

[2] Bickford, J. L.: *Forex Wave Theory-A Technical Analysis for Spot and Futures Currency Traders*, The McGraw-Hill Companies, New York, 2007.

[3] Burgess, G.: *Trading and Investing in the Forex Markets Using Chart Techniques* (Vol. 543), John Wiley and Sons, 2010.

[4] De Villiers, V.: *The Point and Figure Method of Anticipating Stock Price Movements: Complete Theory and Practice*, Stock Market Publications, 1933.

[5] Frost, A. J., and Prechter, R. R.: *Elliott wave principle: key to market behavior*. Elliott Wave International, 2005.

[6] Lo A. W., Mamaysky H., Wang J.: *Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation*. The Journal of Finance, 55(4), 1705-1770, 2000

[7] Logue, D. E., and Sweeney, R. J.: *White noise in imperfect markets: the case of the franc/dollar exchange rates*. The Journal of Finance, 32(3), 761-768, 1977.

[8] Murphy, J. J.: *Technical analysis of the futures markets: A comprehensive guide to trading methods and applications*, Prentice Hall Press, 1986.

[9] Neely C. J., Weller P. A.: *Technical analysis in the foreign exchange market,* Federal Reserve Bank of St. Louis Working Paper No., 2011.

[10] Oberlechner, T.: *The psychology of the foreign exchange market* (Vol. 506). John Wiley and Sons, 2005.

[11] Piasecki K., Stasiak, M. D.: *Exchange rate modeling withuse of relative binary representation*,. Proceedings of 36th International Conference Mathematical Methods in Economics MME 2018 , MatfyzPress, Publishing House of the Faculty of Mathematics and Physics Charles University

[12] Schlossberg, B.: *Technical Analysis of the Currency Market*. John Wiley & Sons Inc, 2006.

[13] Stasiak, M. D.: *Modelling of currency exchange rates using a binary representation,* Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016, Springer, 2017.

[14] Wilder J. W., Jr.: *New concepts in technical trading systems*. Greensboro, NC:Hunter Publishing Company, 1978.

Robust optimization approach in travelling salesman problem

Tereza Nehézová[1]

Abstract. Travelling salesman problem is considered one of the greatest mathematical questions of today. It is defined as finding the shortest Hamiltonian path in a graph while visiting all vertices just once. This paper deals with travelling salesman problem while considering some parts of its mathematical model to be uncertain. In simple travelling salesman problem all the route evaluations are known, but in reality it is a common practice that evaluations of routes can be unknown or uncertain. Uncertainty in optimization models can be handled by using stochastic optimization. In this paper we show how to use robust optimization approach towards modelling uncertainty. The robust approach allows an optimization model to remain relatively simple while finding a robust-optimal solution. It also allows to identify deviations of deterministic values. The way of incorporating of robust aspects in the travelling salesman problem is described in detail and shown on an example in the end.

Keywords: robust optimization, travelling salesman problem, uncertainty

JEL Classification: *C61*
AMS Classification: *90C05, 90C08,90C11*

1 Introduction

Travelling salesman problem (TSP for short) is, how it was described by Applegate [1], problem of finding the cheapest way in set of cities along with the cost of travel between each pair of them while visiting all of them and returning to the starting point. It is meant to find the order in which the cities are visited. This problem is widely known among the mathematicians due to its nature to be easily understood, but till today, there is no general solution, only for specific cases. One of the first and most significant researchers can be named Dantzig, Fulkerson and Johnson [8] and their solution for set of 49 cities, reached by using linear programing. TSP has many usage, not just in transportation, such as in neural networks, logistics and genetics to name just a few. These days there are numerous variations of TSP used to solve specific problems in various areas of usage, for instance Malaguti and Martello [11] solving TSP with pickup, deliveries and draft limits applied in marine transportation. But in most cases all data is known precisely and no uncertainty is considered. In this paper the uncertainty in the data is assumed, the route evaluation to be specific.

Generic optimization model works with the assumption that all parameters are exactly known and are equal to some nominal values. However, this approach does not consider the impact of the uncertainty on the quality and admissibility of the model. Therefore, because of differences in data value and denominations, several model constraints may be disrupted and solutions obtained using nominal data may no longer be optimal or even acceptable. This implies the need to construct models that are immune to data-side uncertainty. The first author to address the issue of robust optimization was Soyster [12], who designed a linear programming optimization model that provides a solution that is resistant to distortion caused by data uncertainty. Brief description of robust optimization, especially in linear programing was made by Ben-Tal, El Ghaoui and Nemirovski [2], issue of robust optimization as a whole was described again by them in [2]. In the last 10 years, there has been a great growth in the application of robust optimization in various sectors. [9].

One of the most - adopted robust optimization approaches is the so-called Γ-scenario set developed by Bertsimas and Sim [5]. This model allows to describe the uncertainty by symmetric ranges of deviation from nominal values and control the number of deviating coefficients by parameter Γ. This approach keeps the model relatively simple, which is an important computational advantage, especially when solving large scale problems. Furthermore, the model still keeps the nature of linear programming. General review of multi-band approach was outlined by Büsing and D'Andreagiovanni [6], followed by new theoretical framework and expansion of the approach again by Büsing and D'Andreagiovanni [7].

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, nehezova@pef.czu.cz

This paper seeks the possibility of using robust approach in travelling salesman problem, which reflects more realistic way of application of TSP in real cases. In this paper a simple way how to reformulate generic travelling salesman problem into its robust counterpart assuming deviation in cost coefficients will be shown.

2 Materials and methods

In this part the way of using robust approach in generic linear optimization model , as well as mathematical description of a travelling salesman problem will be shown.

2.1 Formulation of travelling salesman problem (TSP)

Travelling salesman problem is a well known NP-complete problem. We assume the following general formulation of TSP:

$$z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$$

s.t.

$$\begin{gathered}\sum_{j=1}^{n} x_{ij} = 1, i = 1,2,\dots,n \\ \sum_{i=1}^{n} x_{ij} = 1, j = 1,2,\dots,n \\ u_i - u_j + nx_{ij} \leq n-1, i = 1,2,\dots,n; j = 1,2,\dots,n; i \neq j \\ x_{ij}\epsilon\{0;1\}, i = 1,2,\dots,n; j = 1,2,\dots,n\end{gathered} \tag{1}$$

where n is the number of nodes that are supposed to be visited, $c_{ij}$is the distance between i and j and all the parameters are precisely known, x_{ij} is a bivalent variable that acquires a value of 1 if the connection between the nodes i a j is used, 0 if not, u_{ij} is an auxiliary variable and appropriate constraints prevents the creation of partial cycles.

2.2 Robust optimization approach in linear programing (LP)

In this paragraph it is shown how the generic linear optimization model can be transformed into its robust counterpart, which means to include uncertainty into it. Here, a generic integer linear optimization model is assumed:

$$max \sum_{j=1}^{n} c_j x_j$$

s.t.

$$\begin{gathered}\sum_{j=1}^{n} a_{ij}x_j \leq b_i, i = 1,\dots,m \\ x_{ij}\epsilon\{0;1\}, i = 1,2,\dots,n; j = 1,2,\dots,n\end{gathered} \tag{2}$$

The presence of uncertainty is usually considered as that some of the coefficients a_{ij}, b_j, c_j are not precisely defined. This leads to a new problem:

$$max \sum_{j=1}^{n} (c_j + \delta_j^c)x_j$$

s.t.

(3)

$$\sum_{j=1}^{n}(a_{ij}+\delta_{ij}^{a})x_j \le b_i+\delta_i^{b}, i=1,\dots,m$$

$$x_{ij}\epsilon\{0;1\}, i=1,2,\dots,n; j=1,2,\dots,n$$

The (2) is robust counterpart for (1) where the uncertainty is expressed using deviations for any coefficient if needs to be. Deviations represented as δ_{ij} are assumed to be any nonzero number. Any slight changein the original coefficient value may affect the optimal solution in adverse way and even can cause the infeasibility of how the Ben-Tal, El Ghaoui and Nemirovski [3] illustrated it. Before a robust counterpart of (1) is constructed, there are 4 assumptions proposed by the Bertsimas and Sim [4], that must be held:

1. For each coefficient a_{ij} one is able to define expected (deterministic) value and maximum deviation δ_{ij}^{a} from its deterministic value.
2. Deterministic value a_{ij} then belongs to the symmetric interval $\left[a_{ij}-\delta_{ij}^{a}, a_{ij}+\delta_{ij}^{a}\right]$.
3. The uncertain coefficients are stochasticaly independent random coeficients, each of its own deviation range.
4. For each constraint i it is possible to define maximum number of coeficients Γ_i that will simultaneously deviate from its deterministic value the constaint i.

Then robust counterpart of (1) can be constructed:

$$max \sum_{j=1}^{n} c_j\, x_j$$

s.t.

$$\begin{aligned}
\sum_{j=1}^{n} a_{ij}\, x_{ij} + \Gamma_i z_i + \sum_{j\in U_i} p_{ij} &\le b_i, i=1,\dots,m \\
z_i + p_{ij} &\ge \delta_{ij}^{a} x_j, \quad i=1,\dots,m, \forall j \in U_i \\
z_i &\ge 0, \quad i=1,\dots,m \\
p_{ij} &\ge 0, \quad i=1,\dots,m, \forall j \in U_i \\
x_{ij}\epsilon\{0;1\}, &\quad i=1,2,\dots,n; j=1,2,\dots,n
\end{aligned} \tag{4}$$

where the parameter $\Gamma_i, 0 \le \Gamma_i \le |U_i|, i=1,\dots,m$ can be used to control uncertainty for corresponding constraint i, value of Γ sets th maximum number of lefthanded coeficients a_{ij} that are assumed to deviate by a maximum of δ_{ij}^{a}, p_{ij} is an auxiliary variable for each a_{ij} that is considered uncertain, z_i is another auxiliary variable merely preserving a relationship between the first and second constraint and U_i indicates a set of indices j of those a_{ij} which are actually considered to be uncertain.

3 Results and discussion

When there are defined both the robust optimization counterpart of LP and travelling salesman problem, robust approach towards TSP can be defined. As it has been shown in (2), the coefficients which can be uncertain are cost coefficients, left-handed and right-handed coefficients. In travelling salesman problem left-handed coefficients are bivalent constrains indicating whether the given link x_{ij} was realized, right-handed are necessarily equal to one and ensure that each node is visited just once. Cost coefficients if representing route time can perform uncertain behavior. In real application route time can be affected by traffic situation, such as traffic jams or unforeseen events on the road, that would cause cost coefficients, e. g. time, to grow.

3.1 Formulation robust counterpart of travelling salesman problem

The uncertainty in TSP will be laid on objective function coefficients c_{ij}. The deviation of the expected value of c_{ij} has to be included to the objective function as δ_{ij}^{c}. We assume objective function including uncertainty in c_{ij}:

$$min \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} + \delta_{ij}^{c}) x_{ij}$$

which can be represented also as:

$$min\ E$$

s.t.

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} + \delta_{ij}^{c}) x_{ij} \leq E$$

Now robust counterpart of TSP can be done:

$$\begin{gathered} min\ E \\ \text{s.t.} \\ \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} + \Gamma z + \sum_{(i,j) \in U_i} p_{ij} \leq E \\ z_i + p_{ij} \geq \delta_{ij}^{c} x_j, \forall (i,j) \in U_i \\ \sum_{j=1}^{n} x_{ij} = 1, i = 1,2, \dots, m \\ \sum_{i=1}^{m} x_{ij} = 1, j = 1,2, \dots, n \\ z \geq 0 \\ p_{ij} \geq 0, \forall (i,j) \in U_i \\ u_i - u_j + n x_{ij} \leq n - 1, i = 1,2, \dots, n; j = 1,2, \dots, n; i \neq j \\ x_{ij} \epsilon \{0; 1\}, i = 1,2, \dots, n; j = 1,2, \dots, n \end{gathered} \tag{5}$$

where δ_{ij}^{c} is maximal deviation of c_{ij}, parameter $\Gamma_i, 0 \leq \Gamma_i \leq |U_i|$ allows one to control the uncertainty for the selected constraint i, parameter value Γ indicates the maximum number of left-hand coefficients c_{ij}, which are expected to deviate by no more than δ_{ij}^{c}, p_{ij} and z has the same meaning as in (4). U_i indicates the set of indices j of those a_{ij} for which the deviation is actualy considered.

3.2 Practical example

For illustration how it is possible to implement the uncertainty into the travelling salesman problem, in the cost coeficients to be specific, and how the change of initial parameters influences the objective value, we will assume a case with 10 nodes that has to be visited. Matrix of distances is not symmetrical and objective cost is assumed in minutes. Solving that model as (5) if $\Gamma = 0$, we assume that no c_{ij} will deviate and we receive a solution with objective value 162,5 and order of nodes that can be seen in following table:

x_{ij}	x_{15}	x_{53}	x_{310}	x_{108}	x_{89}	x_{96}	x_{62}	x_{27}	x_{74}	x_{41}	Objective value
c_{ij}	15,983	19,883	12,55	23,567	30,883	7,75	6,767	7,167	20,067	17,883	162,5

Table 1 TSP solution

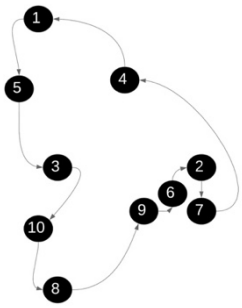


Figure 1 Display of route

4 scenarios of robust solution will be inspected in dependence on parameter Γ that represents the highest number of coefficients c_{ij} that will deviate by δ_{ij}^{c} in the objective function. That means if $\Gamma = 10$, c_{ij} will deviate from its deterministic value, while it is not given particular c_{ij}, that would be deviating. More robust solutions are given by higher value of Γ, because more deviations of cost coefficients will be assumed and that will lead to the increase in objective value. Scenarios are considered with maximum 25% deviation of c_{ij}.

If we change Γ, for example $\Gamma = 3$, it expresses that at most 3 coefficients will deviate by its deterministic value. According to [2], even slight change of cost coefficients can affect the optimal solution. This case was shown in the case of transportation problem in [9]. In TSP, change does not have to lead to the change of the order of nodes, but the model can show specific joints, that are appearing in scenarios multiple times with deviations from its deterministic value, and may be problematic be planned. In all the tested scenarios, the order of nodes remains the same as in the case without any deviation, but particular joints appeared more often with deviation from their deterministic value than others. The appearance of deviation in tested scenarios is shown (Table 2).

Value of Γ	Deviated indicies										Objective value
1											170,2208
2	p_{89}										176,1125
3	p_{108}	p_{89}									181,1129
4	p_{108}	p_{89}	p_{74}								186,0999
5	p_{53}	p_{108}	p_{89}	p_{74}							193,8208
6	p_{15}	p_{53}	p_{89}	p_{74}	p_{41}						194,5666
7	p_{15}	p_{53}	p_{108}	p_{89}	p_{74}	p_{41}					197,7041
8	p_{15}	p_{53}	p_{310}	p_{108}	p_{89}	p_{74}	p_{41}				199,6416
9	p_{15}	p_{53}	p_{310}	p_{108}	p_{89}	p_{96}	p_{74}	p_{41}			201,4333
10	p_{15}	p_{53}	p_{310}	p_{108}	p_{89}	p_{96}	p_{27}	p_{74}	p_{41}		203,1249
11	p_{15}	p_{53}	p_{310}	p_{108}	p_{89}	p_{96}	p_{62}	p_{27}	p_{74}	p_{41}	203,125

Table 2 Different robust solutions

It can be seen (Table 2), that the most problematic variable is p_{89} which represent route between nodes 8 and 9 and it appear except for $\Gamma = 1$ in every scenario. Other problematic parameters are then p_{108} and p_{74}. Rise of the objective value depends on the number of deviating coefficients and correspond to the situation, when coefficients deviate by maximum possible value of δ_{ij}^{c}. Moreover, more than $\Gamma = 11$ does not show any change in objective value or order of the nodes, since the TSP is a case of bivalent programming and the nodes that are not used have $x_{ij} = 0$.

4 Conclusion

The travelling salesman problem robust counterpart was constructed using Γ - robustness approach and has shown how uncertainty can be treated without using stochastic programming. Uncertain conditions are commonly appearing in transportation, mostly because of traffic conditions on routes. The robust solution as used in this paper can show the problematic part of the routes, that can appear as deviated multiple times and also show

total transportation time of the worst scenarios. This aspect can be useful for decision maker to know in advance what to expect and to refine his choices accordingly. The changes of cost coefficients may also lead to the change of the optimal order of nodes. The approach used in this paper is also less demanding than the comparison and evaluation of all possible cases with different values. Robust optimization model still retains the nature of linear programming and thus is easy computationally tracable, which is a great advantage in case of already complex model like TSP.

References

[1] Applegate, David L., *The traveling salesman problem: a computational study*. Princeton: Princeton University Press, 2006

[2] Ben-Tal, A. and Nemirovski, A.: Robust solutions of uncertain linear programs. Operations Research Letters **25** (1999), 1-13.

[3] Ben-Tal, A., El Ghaoui A.L. and Nemirovski A.: Robust optimization. Princeton University Press, 2009.

[4] Bertsimas, D. and Sim, M.: Robust discrete optimization and network flows. Mathematical Programming **98** (2003), 49-71.

[5] Bertsimas, D. and Sim, M.: The price of robustness. Operations Research (52) (2004), 35-53.

[6] Büsing, C. and D'Andreagiovanni, F.: Robust optimization under Multi-band uncertainty – Part I: Theory, 2013.

[7] Büsing, C. and D'Andreagiovanni, F.: A New Theoretical Framework for Robust Optimization under Multi-band Uncertainty. In: Operations Research Proceedings, 2014, 115-121.

[8] Dantzig, G., R. Fulkerson a S. Johnson, Solution of a Large-Scale Traveling-Salesman Problem. *Journal of the Operations Research Society of America*. 2(4) (1954), 394-410.

[9] Gabrel, Virginie, Cécile Murat a Aurélie Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*. 2014, 235(3), 471-483

[10] Hlavatý, Robert a Helena Brožová, 2017. Mathematical Methods in Economics 2017: Robust optimization approach in transportation problem.

[11] Malaguti, Enrico, Silvano Martello a Alberto Santini, 2018. The traveling salesman problem with pickups, deliveries, and draft limits. *Omega* 24. 1. 2017, (74**)**, 50-58

[12] Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research, 21, (1973), 1154-1157.

Estimating the environmental efficiency of European cross-countries using a non-radial general directional distance function

Lucie Chytilova[1], Jana Hanclova[2]

Abstract. This article deals with the evaluation of the environmental efficiency of 29 EU countries in the time period from 2000 to 2016. The classic version of the additive data envelopment (DEA) model is modified by the non-radial general directional distance function for distance measurement. The proposed model makes it possible to reduce the input variables (labor, capital, energy), to reduce the undesired output variables (greenhouse gas, GHG) and to increase desired output variables (gross domestic product, GDP) in the same time and non-proportionally. The results show the development of the average value of scaling factor beta, which measures the inefficiency, which has been decreasing from the average since 2004. Reducing inefficiency of work and reducing GHG emissions are a significant contributor to this. Also, this is related to the land mark environmental policy of reducing GHG emissions after 2005 in line with the Kyoto Protocol. Throughout the period of 17 years, according to GDP-GHG index, the environmental efficiency were 14 EU countries with best practice, but on the other hand, the least environmental efficient countries were Latvia, Poland and Romania.

Keywords: Data envelopment analysis, generalized directional distance function, undesirable output, non-radial measures, non-proportionally measures, EU countries, environmental efficiency, GHG emissions.

JEL Classification: C61, O47, Q51
AMS Classification: 90C05, 90C90

1 Introduction

The European Union has created the European Union Emissions Trading Scheme (EU ETS), and land mark environmental policy was launched under article [1] in January 2005 by reducing Greenhouse Gas (GHG) emissions. The whole process is divided into three phases in accordance with the Kyoto Protocol. Phase I took place between years 2005 and 2007. The second phase was from 2008 to 2012 and the goal was to reduce emissions from 2008 to 2020 by 20% and increase the share of renewables in the energy mix to 20%. Thus, the third phase (2013-2020) is currently underway with the aim of centralizing National Allowance Plans, harmonizing its operation and extending it to other national economies.

Utilization of Data envelopment analysis (DEA) in energy or environmental studies is mainly seen in review articles by Zhou et al. [13] and Mardani et al. [7]. It turns out that the number of articles applying DEA models with desirable and undesirable output variables at regional or national level has a growing tendency. Ramamathan in [8] has presented the use of the classic radial DEA models with three variables representing CO2 emissions, energy consumption and economic activity for 64 countries. He concluded that Central European countries (Poland, Czech Republic, Romania) and South Africa are least efficient. Fare et al. in article [4] introduced the index of environmental performance and this index is based on efficiency measure using distance functions. Their application was for OECD countries in 1990. Other authors Zhou et al. in a study [14] developed two slack-based efficiency measures for modeling economic-environmental DEA performance for 30 OECD countries for time period 1998-2000. The publication [9] was focused on eco-efficiency in 26 European countries using DEA model. Results and outcomes of this study indicated some countries such Ireland, Hungary, Slovakia and Portugal were most efficient and some least efficient (Italy, Denmark, Bulgaria and Romania). In article [3], authors tried also the special DEA model for the EU countries.

[1] VSB-Technical University of Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 2416/33, 702 00 Ostrava, Czech Republic, lucie.chytilova@vsb.cz
[2] VSB-Technical University of Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 2416/33, 702 00 Ostrava, Czech Republic, jana.hanclova@vsb.cz

Zhou et al. has presented in article [15] the new non-radial DEA model based on the concept of environmental performance of OECD countries. Other authors as Gavurova et al. [5] also used the non-radial and non-oriented SBM model with undesirable output variable under the condition of the return to scale in OECD countries during time period 1995-2014. Authors concluded that the lowest efficiency has been in the Czech republic, Hungary, Poland , and Slovakia. The Lithuania had the highest progress from 10.17% to 26.22% in 2014. The article by Hsieh et al. [6] examined the dynamic DEA for evaluation of energy environmental efficiency of 28 countries in the European countries during the period 2006-2013. The input variables were labor, capital and energy. Undesirable output variable were greenhouse gas emissions and sulfur oxide emissions, and desirable output was the gross domestic product. The carry-over variable was gross capital formation. Results of dynamic DEA model show that the environment efficiency has improved for the inefficient countries during the years 2009-2012. It is also confirmed that the desirable output variables and undesirable output variables have had a significant effect on environmental efficiency.

The aim of this article is to evaluate the GDP-environmental efficiency for 29 European countries in time period from 2000 to 2016 through the non-radial DEA model with non-proportional distance function. We also investigate the effect of input and output variables based on the development of GDP-GHG environmental efficiency.

The rest of the paper is organized as follows. Section 2 introduces methodology on non-radial DEA with non-proportional directional distance function. Section 3 describes the data, and presents the results with discussion. Section 4 includes conclusions.

2 Methodology

2.1 Environmental production technology

To model the production of N national economies ($j = 1, 2, ..., N$) we will consider a conceptual model with a joint production framework of R desirable output variables $\mathbf{y}$ and S undesirable output variables $\mathbf{b}$ (for example: greenhouse gas emissions; the designation $\mathbf{b}$ is based on the "bad" outputs). We also assume the group I multiple input variables $\mathbf{x}$. The multiple-output environmental production technology with emphasis on input-specific technology can be described as $T_1(\mathbf{x}) = \{(\mathbf{y}, \mathbf{b}) : \mathbf{x} \text{ can produce } (\mathbf{y}, \mathbf{b})\}$ according to [10]. $T_1(\mathbf{x})$ is called an *environmental output variable* assuming the following two axioms which are fulfilled by [10]:

1. Weak disposability of output variables: The set of all proportional reductions of output variables are feasible for a given fixed level of input variables $\mathbf{x}$, i.e. if $(\mathbf{y}, \mathbf{b}) \in T_1(\mathbf{x})$ and $0 \leq \theta \leq 1$, then $(\theta\mathbf{y}, \theta\mathbf{b}) \in T_1(\mathbf{x})$.
2. Null-joint: The undesirable output variables are *null-joint* with the desirable output, i.e. if $(\mathbf{y}, \mathbf{b}) \in T_1(\mathbf{x})$ and $\mathbf{b} = 0$, then $\mathbf{y} = 0$.

2.2 Non-radial and non-proportional generalized directional function

Directional functional distance can be used to measure environmental production efficiency. To increase the desirable output and reduce the undesirable output assuming fixed inputs, Chung et al. in the publication [2] have introduced *the directional output distance function* (DDF) and the joint production of desirable output and undesirable output b:

$$\overrightarrow{D}_{\mathbf{T_1}}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^y, \mathbf{g}^b) = \sup\{\beta | (\mathbf{y}, \mathbf{b}) + \beta(\mathbf{g}^y, \mathbf{g}^b) \in T_1(\mathbf{x})\}, \tag{1}$$

where the nonzero vector $(\mathbf{g}^y, \mathbf{g}^b)$ is the *direction vector* and β it expresses the intensity of increased desirable production and at the same time decreases undesireble production and is referred to scaling factor. $\overrightarrow{D}_{\mathbf{T_1}}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^y, \mathbf{g}^b) \geq 0$ if and only if $(\mathbf{y}, \mathbf{b}) \in T_1(x)$. This DDF function moves the joint production $(\mathbf{y}, \mathbf{b})$ along the direction $(\mathbf{g}^y, \mathbf{g}^b)$ to place it on the production frontier using following model (2):

$$\begin{array}{lll} & z_{(2)} = \max \beta & \\ \text{s.t.} & \sum_{i=1}^{I} \lambda_j \mathbf{x}_{ij} \leq \mathbf{x}_{io} & i = 1, 2, ..., I \\ & \sum_{r=1}^{R} \lambda_j \mathbf{y}_{rj} \geq \mathbf{y}_{ro} + \beta g_r^y & r = 1, 2, ..., R \\ & \sum_{s=1}^{S} \lambda_j \mathbf{b}_{sj} = \mathbf{b}_{so} + \beta g_s^b & s = 1, 2, ..., S \\ & \sum_{j=1}^{N} \lambda_j = 1 & \lambda_j \geq 0. \end{array} \tag{2}$$

Model (2) for $(\mathbf{g}^y, \mathbf{g}^b) = (\mathbf{1}, -\mathbf{1})$ is a classic output oriented DEA model with undesirable output variables under variable return to scale (VRS). This model *proportionally* reduces undesirable output variables and increases desired input variables. Zhou et al. showed in article [13] that radial method for measuring efficiency may be overestimated if we have non-zero slack variables. In accordance with the publication [9], he proposed a **non-radial**

model where the direction vector is $\mathbf{g} = (\mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b)' = (-\mathbf{x_o}, \mathbf{y_o}, -\mathbf{b_o})$. Assuming a **non-proportional** change in input and output variables using a scaling vector $\beta = (\beta^x, \beta^y, \beta^b)'$. We can formulate a model (3) for determining environmental efficiency:

$$
\begin{aligned}
& z_{(3)} = \overrightarrow{D}_{\mathbf{T_2}}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b) = \max\{\mathbf{w}^{x\prime} \cdot \beta^x + \mathbf{w}^{\mathbf{y}\prime} \cdot \beta^{\mathbf{y}} + \mathbf{w}^{\mathbf{b}\prime} \cdot \beta^{\mathbf{b}}\} = \beta^* \\
\text{s.t.} \quad & \textstyle\sum_{i=1}^{I} \lambda_j \mathbf{x}_{ij} \leq (1 - \beta_i^x)\mathbf{x}_{io} && i = 1, 2, ..., I \\
& \textstyle\sum_{r=1}^{R} \lambda_j \mathbf{y}_{rj} \geq (1 + \beta_r^y)\mathbf{y}_{ro} && r = 1, 2, ..., R \\
& \textstyle\sum_{s=1}^{S} \lambda_j \mathbf{b}_{sj} = (1 - \beta_s^b)\mathbf{b}_{so} && s = 1, 2, ..., S \\
& \textstyle\sum_{j=1}^{N} \lambda_j = 1 && \lambda_j \geq 0 \\
& \beta_i^x, \beta_r^y, \beta_s^b \geq 0 && \forall\, i, r, s,
\end{aligned}
\tag{3}
$$

where vector $\mathbf{w}' = (\mathbf{w^x}, \mathbf{w^y}, \mathbf{w^b})$ is the normalized weight vector and we assume that the weight of all input, desirable and un desirable output variables are gradually $1/3$. Model (3) is similar to the additive DEA model in the sense that both attempt to identify potential slacks in input and output variables as much as possible. The non-radial directional distance function is based on T_2:

$$
\overrightarrow{D}_{T_2}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b) = \sup\{\mathbf{w}'\beta : |(\mathbf{x}, \mathbf{y}, \mathbf{b}) + \mathbf{g} \cdot \ diag\ (\beta) \in T_2\}. \tag{4}
$$

To evaluate the output-environmental *y-b performance index and efficiency* we use Zhou et al. (2012) $YBPI$ Index:

$$
YBPI = \frac{\frac{1}{2}[(1 - \frac{1}{I}\sum_{i=1}^{I} \beta_i^{x*}) + \frac{1}{S}(1 - \sum_{s=1}^{S} \beta_s^{b*})]}{\frac{1}{3}(1 + \frac{1}{R}\sum_{r=1}^{R} \beta_r^{y*})} = \frac{1 - \frac{1}{2}((\frac{1}{I}\sum_{i=1}^{I} \beta_i^{x*}) + \frac{1}{S}(\sum_{s=1}^{S} \beta_s^{b*}))}{\frac{1}{3}(1 + \frac{1}{R}\sum_{r=1}^{R} \beta_r^{y*})}. \tag{5}
$$

The numerator in (5) represents the average proportion by which the input and desirable output variables can be reduced, while the denominator expresses the degree to which desirable output variable can be increased. Index $YBPI$ is very similar to the slacks-based efficiency measure such as unit invariance and monotonicity. Index $YBPI$ is standardized between 0 and 1, $YBPI = 1$ means that the country is located at the frontier of best practice.

3 Empirical study

The empirical study is devoted to exploring output-environmental efficiency for selected European countries in time period 2000-2016. Estimates are based on the non-proportional DEA model (3) with non-radial general directional distance function.

I_O	ID	variable	unit	definition
inputs	**C**	Capital	Chain linked volumes (2010), million euro	Gross capital formation
	L	Labor	Thousand hours worked	Total employment domestic concept
	E	Energy	Thousand tones of oil equivalent (TOE)	Final consumption - energy use
desirable output	**Y**	GDP	Chain linked volumes (2010), million euro	Gross domestic product
undesirable output	**B**	GHG	Thousand tonnes in CO_2 equivalent (TOE)	Greenhouse gases (CO2, N2O, CH42 HFC, PFC, SF6, NF3)

Table 1 Description of variables in the DEA model [Source: Eurostat database]

The VRS model is used in this empirical study because, although all countries are from the European Union, the environment in these countries is at a different level as well as its support.

3.1 Data

Analysis of output-environmental efficiency was performed for 29 European countries (Belgium, Bulgaria, the Czech Republic, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Finland, Sweden, United Kingdom,

Norway). Based on expert articles and using Eurostat[1] were as variables define 3 input variables, 1 desirable output variable and 1 desirable output variable. Table 1 presents a description of the variables which are used in the DEA model. The trend in the average value of Greenhouse gases has been declining slightly, with an annual average of 1 878 thousand tones in CO_2 equivalent. The standard deviation (SD) statistics show that for the labor and GDP variables there is a slight increase in SD and, conversely, energy consumption decreases and the SD is stable for the GHG variable.

3.2 Results and discussions

GAMS software was used to optimize the model (3). Development of average values of optimal beta coefficients $\beta_c^{x*}(BETA)$, $\beta_L^{x*}(BETALCX)$, $\beta_E^{x*}(BETALX)$, $\beta_c^{x*}(BETAEX)$, $\beta_{GDP}^{y*}(BETAY)$ and $\beta_{GHG}^{b*}(BETAB)$ are shown in Figure 1 for 29 countries in the time period 2000-2016.

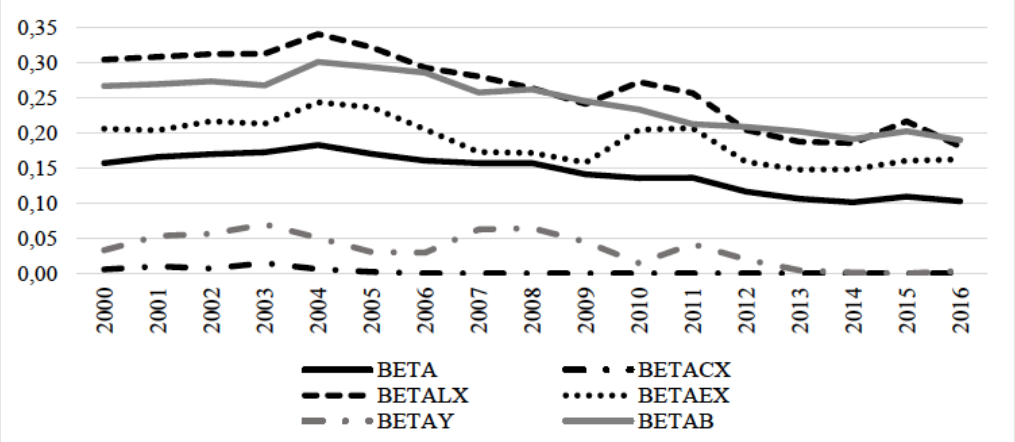


Figure 1 Development of average values of optimal beta coefficients in years 2000-2016

The results show that the overall inefficiency expressed by the optimal value of the purpose function β^* (BETA) has been decreasing since 2004. The main factor of inefficiency is the need to reduce labor input variable. This trend started from the year 2004, when beta LX declines from 0.34 in 2004 to 0.18 in 2019 with a slightly decreasing deviation. Reducing the inefficiency of greenhouse gases GHG is also necessary to reduce for the inefficiency, and since 2004 BETAB has been reduced from 0.30 to 0.19 in 2016, but lower than BETALX. The average inefficiency of BETAEX raged in 2004-2005 and also in 2010-2011. Decreases after year 2004, especially after 2005, are related to the signing of the Kyoto Protocol document and the launch of Phase I with emission reduction. The level of capital entry is close to zero (BETACX) and therefore there is no need to reduce this input variable on average. Increasing output variable to reduce output-environmental inefficiency was recorded only in the time period 200-2013 with peaks in 2003, 2008 and 2011. The last two years are probably related to the financial and economic crisis in two waves.

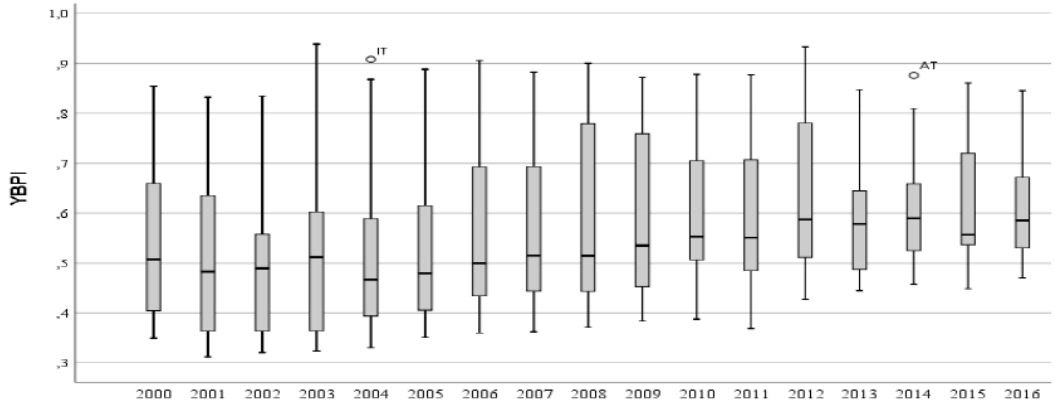


Figure 2 Development of GDP_B efficiency index in years 2000-2016

[1] https://ec.europa.eu/eurostat/data/database

Equation (5) was used to determine the GDP-GHG environmental performance index $YBPI$. Figure 2 presents the box-plots of this index in time period 2000-2016 for a group of inefficient countries. The analysis of the detailed results showed that 14 countries - according to $YBPI$ - were on the efficient border with the best practice across the whole 17 years - Belgium, Bulgaria, the Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Luxembourg, Malta Kingdom, Norway. On the other hand, Latvia, Poland and Romania were the least effective in the entire period. Latvia has a high input of labor factor and on the second place is the high energy consumption with a low GHG emission reduction rate. In the case of Poland, it is necessary to reduce GHG emissions as well as work intensity and to reduce energy consumption in the second place. Romania should focus on reduced work intensity, further by reducing emissions and reducing energy consumption in a non-aligned fashion.

It is also possible to observe the growing trend in the development of $YBPI$ since 2004, ie the improvement of the GDP-environmental efficiency median and the kvantil margin of $YBPI$ reduction, especially after 2013.

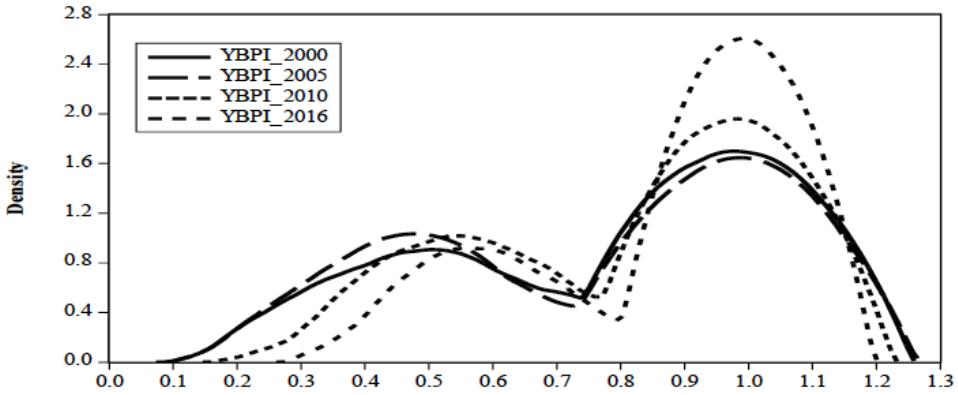


Figure 3 Development of GDP-GHG efficiency index in years 2000-2016

Figure 3 compares GDP-GHG performance to the performance of the European countries under review in 2000, 2005, 2010 and 2016. GDP-GHG performance is more likely to be equal to 1, ie more countries are on efficient border. For a group of inefficient countries according to the $YBPI$ index, the probability of the top of a group of inefficient countries shifts to a higher level of $YBPI$, ie the efficiency level increases.

4 Conclusion

The main results of the measurement and evaluation of environmental efficiency development are connected with the modification of the classical DEA model, especially in two directions - the introduction of non-radial general directional distance function for measuring the distance of inefficient units from the best practice boundary. And also the possibility of monitoring non-proportional reduction of labor and capital or energy (input variables), reducing GHG emissions and simultaneously increasing GDP of the desired output variables of the European Union's economies in time period 2000-2016.

The obtained conclusions are supported by a further scientific literature, for example by Robain et al. [9] and Yao et al. [11]. It indicates that after 2004 the average deviation from the effective boundary, measured by the beta coefficient, is decreasing. This positive development is mainly due to the reduction of the input variable factor of worked hours, which is due to the improvement of technological progress in national economies. The second important factor is the reduction of undesirable emissions expressed by greenhouse gas, which in turn is related to the reduction of emissions by the Kyoto Protocol and thus the reduction of emissions under the Kyoto Protocol since Phase I since 2005. Another positive direction is the slight increase in the number of environmental efficient countries in the years. 2000-2014 and the stability of 14 countries, effective throughout the period under review. On the other hand, it is indispensable to address the relatively low environmental efficiency in Latvia, Poland and Romania, with the greatest problem being high work intensity, the need to reduce GHG emissions and reduce energy consumption. Latvia has been slightly improving this undesirable trend since 2011, but in the case of Poland and Romania, the unwanted position is stagnating.

Further research in this area will be focused on the possibilities of monitoring the effect of non-proportional emission reduction for different GHG components, exploring the causes of environmental inefficiency by means of panel regression, and exploring the environmental dynamics efficiency using a modified network DEA model.

Acknowledgements

This article was supported by European Social Fund within the project CZ.1.07/2.3.00/20.0296 and also through the Czech Science Foundation GACR 19-13946S.

References

[1] Chevallier, J. (2012). *Econometric analysis of carbon markets: the European Union emissions trading scheme and the clean development mechanism*. Dordrecht: Springer.

[2] Chung, Y. H., Fro, R. & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51(3), 229–240.

[3] Chytilova, L. & Zapletal, F. (2015). *An efficiency evaluation of emission allowances for EU countries with the use of the zero sum game data envelopment analysis*. In proceedings of international conference METAL 2015. Brno.

[4] Fare, R. & Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation: Comment. *European Journal of Operational Research*,157, 242–245.

[5] Gavurova, B., Kocisova, K., Behun, M. & Tarhanicova, M. (2018). Environmental performance in OECD countries: A non-radial DEA approach. *Acta Montanistica Slovaca*, 23(2), 206–215.

[6] Hsieh, J. Ch., Lu, Ch . Ch., Li, Y., Chiu, Y. H. & Xu, Y. S. (2019). Environmental Assessment of European Union Countrie, *Energies*, 12(295), 1–8.

[7] Mardani, A., Zavadskas, E. K., Streimikiene, D., Jusoh, A. & Khoshnoudi, M. (2017). A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency. *Renewable and Sustainable Energy Reviews*, 70, 1298–1322.

[8] Ramanathan, R. (2002). Combing indicators of energy consumption and CO2 emissions: A cross-country comparison. International Journal of Global Energy, 17, 214–227.

[9] Robaina-Alves, M., Moutinho, V. & Macedo, P. (2015). A new frontier approach to model the eco-efficiency in European countries. *Journal of Cleaner Production*, 103, 562–573.

[10] Toloo, M. & Hanclova, J. (2019). Multi-valued measures in DEA in the presence of undesirable outputs. *Omega*, in press.

[11] Yao, Y., Han X. & Wang C. (2019). Can constraint targets facilitate industrial green production performance in China? Energy-saving target vs emission-reduction target, *Journal of Cleaner Production*, 209(2019), 862–875.

[12] Zhou, P., Ang, B. W. & Poh, K. L. (2008). A survey of data envelopment analysis in energy and environmental studies. *European Journal of Operational Research*, 189(1), 1–18.

[13] Zhou, P., Ang, B. W. & Wang, H. (2012). Energy and CO2 emission performance in electricity generation: A non-radial directional distance function approach. *European Journal of Operational Research*, 221, 625–635.

[14] Zhou, P., Ang, B.W. & Poh, K. L. (2006). Slacks-based efficiency measures for modelling environmental performance. *Ecological Economics*, 60, 111–118.

[15] Zhou, P., Poh, K.L. & Ang, B.W. (2007). A non-radial DEA approach to measuring environmental performance. *European Journal of Operational Research*, 178, 1–9.

Electricity Consumption Cost for Households in the Czech Republic Based on the High and Low Tariff Rates Ratio – Optimization Model

Martina Kuncová[1], Jana Sekničková[2]

Abstract. The electricity market in the Czech Republic started its transformation and liberalization in the year 2002. Since 2006 also households can choose its electricity supplier and the product offered by a supplier. Although Energy Regulatory Office (ERU) offers the online calculator to compare the electricity consumption cost of the given household, the choice of products is still wide, and the conditions and formula for the electricity consumption cost are hardly understandable. The online calculator assumes an estimation (kWh) for the consumption in the high and low tariff rate. As in the real-life situation, it is just a rough estimate, the final product choice could be wrong, and the expected annual electricity consumption costs might be different from the real ones. Based on our previous analyses of the D25d distribution rate best product selection (for the given household), the optimization model is created to find for what ratio of the high and low tariff rates the given product is the cheapest one. Data are taken from the ERU for the years 2017 and 2018, where 60 products of 29 suppliers are offered for each year and each of the 3 distribution regions (PRE, CEZ, E.ON).

Keywords: Electricity consumption, Dual tariff rates, Households, Costs, Optimization Model.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

One of the commodities that significantly affects each of us is electricity. Only when the electricity supply is interrupted, we realize how big the impact of electricity on our lives is. Electricity makes it easy for us to work and study, it allows us to relax, keep food, etc. Each of us use it for shining, using electronic devices such as televisions, computers, mobile phones and electrical appliances like refrigerators, freezers, stoves and other home appliances and also most current companies would not be able to operate without electricity. No one is surprised that household electricity demand is not declining (Figure 1).

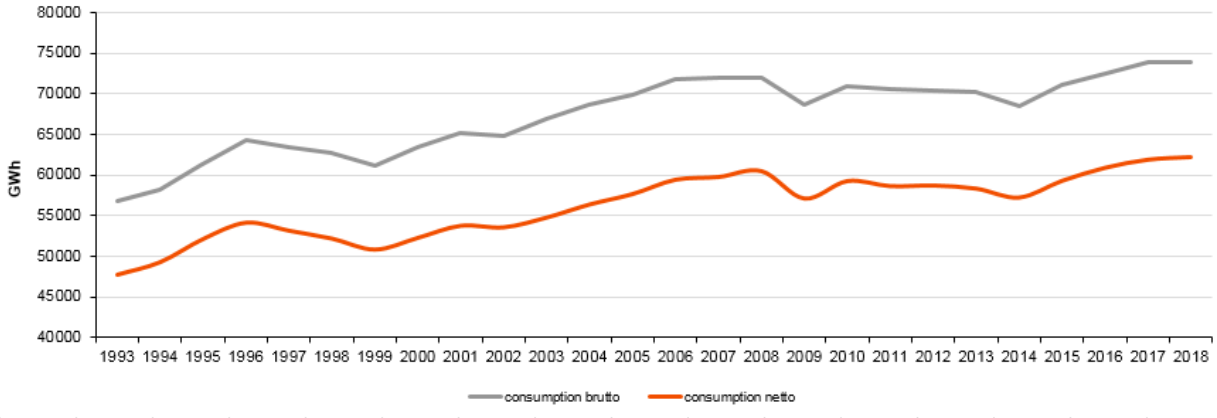


Figure 1: Time series of electricity consumption, source: [2]

[1] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, kuncovam@vse.cz.
[2] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, jana.seknickova@vse.cz.

In 2006, the Czech electricity market for households underwent a major transformation and individual households (as well as companies) can choose their electricity supplier. Logically, households prefer suppliers that provide them with cheaper electricity. The cost of household electricity is also affected by the distributor, the Energy Regulatory Office (ERU) and the market operator (OTE) [12]. Three distributors operate on the Czech market, each managing a different area. Pražská energetika (PRE) covers the capital city of Prague, E.ON Česká republika, s. r. o (E.ON) administers the whole of the South Bohemian and South Moravian regions and the Vysočina Region (except the Havlíčkův Brod District), the Zlín Region (except Vsetín District) and district Prostějov in the Olomouc Region. The remaining regions of the Czech Republic are assigned to ČEZ, a. s. (CEZ) – see Figure 2.

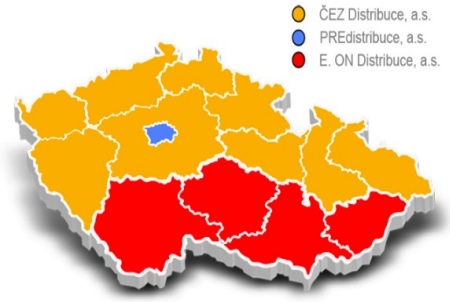


Figure 2: Map of electricity distributors, source: [4]

The number of electricity suppliers on the Czech market is changing every year, and most of them offer several possible products for households. The complete list of products can be obtained with the help of various web calculators. Products then differ in terms and conditions. The final decision depends on the contract conditions but mainly on the prices. This makes it difficult for consumers to select the most suitable supplier and product. Since it is essentially an optimization task, different approaches can be used for modelling like in [14] or [6]. We can use an optimization model from a multi-criteria perspective point of view (see [2] or [12]) or a simulation model as in [6] and [8] or a classical one-criterion optimization view based on costs minimization [9]. This approach will also be used in this article. We continue in the analyses of the suppliers' selection for the households in the Czech electricity market with respect to the level of consumption and annual cost influenced by the suppliers' prices (when the distribution rate D25d is used). This model depends on the suppliers' and distributors' prices, total electricity consumption and also on consumption in the low and high tariff. In [10], we analyzed how the actual electricity consumption of a selected household at D25d rate can move so that the product of the selected supplier is most advantageous (cheapest) for the household. The main objective of our further analysis presented in this article is to decide how the ratio of the low and high tariff of actual electricity consumption can change so that the product of the selected supplier remains the best for a consumer. The results will be then compared with the conclusions of previous research [7], [9] and [10].

2 The situation in the Czech Electricity Market

The liberalization of the electricity market in EU started in 1999 [11], and the Czech Republic joined in 2002 when the companies could enter the market, but since 2006 also the households can choose the electricity supplier on the retail market [1]. Several major entities operate on the Czech electricity market. The electricity market operator (OTE) predicts electricity consumption across the market and conducts market behavior analysis [12]; the Energy Regulatory Office (ERU) then regulates electricity transmission and distribution prices [5]. Distributors are in charge of managing the distribution network and distributing electricity to consumers. Suppliers then offer households their tariffs, which by their conditions create the final price of distributed and supplied energy. OTE and ERUs affect all customers in the same way and the distributor is uniquely determined by the location of the end customer. Thus, none of these entities is choosing by a household. On the contrary, a household can choose the most advantageous product for itself offered by the relevant supplier. As mentioned above, the number of suppliers and their products is quite wide. In recent years, numbers have stabilized on about 60 products from about 30 suppliers [5], but still there are several extinction of suppliers or several new companies entered this market.

The choice of the most suitable product is mainly influenced by the amount of electricity consumed, the agreed rate (determined by the distributor based on the conditions and type of consumption and the circuit breaker) and the prices for consumed electricity. The final price is then influenced by fixed fees and taxes. In general, the final price can be divided into two components. The first one is the fee for the transport of electricity to the end customer and the ERU decides on its amount annually [5]. It includes:

- monthly lease for the circuit breaker,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- price per system services,
- price for the support of the renewable energy purchase,
- charges for the electricity market operator,
- electricity ecological tax (28,30 CZK per 1 MWh).

The second part of the total price is given by the electricity supplier. It covers:

- fixed monthly fee for the selected product,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT).

The final price is increased by VAT, that is 21%.

3 Data and Methods

The number of products offered to the Czech households changes every year and also the number of suppliers is changing. In 2017 based on [5] 29 companies offering 59 or 60 products (company E.ON as a supplier offers 1 product more for the distribution area of PRE and CEZ) were mentioned, for 2018 we see also 29 companies but detailed analysis shows that 3 companies left and 3 entered the market. As suppliers', distributors' and ERUs' prices change every year and they differ nearly in every distribution area and every supplier's product, it is not easy for the household to monitor all the changes and choose the suitable product although ERU tries to simplify the situation by its calculator [5]. For our analysis, all offered products in the year 2017 and 2018 (for D25d distribution rate) were selected. It is evident that the suppliers' average prices and monthly fees are increasing in 2018 compared to 2017 [10]. The biggest change can be seen in low tariff average prices per 1 MWh, where the year-on-year change was up to 15%, followed by the suppliers' monthly fee with the growth of 13%.

In the previous analysis [7], [8], [9], [10] and [12] we compared products for the selected household with the D25d rate. This tariff rate is given to household when the electricity is also used for the accumulative heating and hot water heating for lower and middle yearly offtake with operative management of the validity period of the low tariff for 8 hours. It is the so-called dual tariff rate as it has 2 periods (high tariff, low tariff) during the day. The remaining 16 hours fall into a high tariff. The situation for a household with a circuit breaker from 3x20A to 3x25A and with the fixed power consumption of 10 MWh was analyzed. The analysis was performed, assuming that 45% of the energy is collected at a high rate and 55% at a low rate. In this article, we use the optimization models to change this high and low tariff rate ratio to find the cheapest products.

The choice of a suitable product in terms of single-criterion optimization should be made by the customer based on the total cost of distribution and supply of electricity. The final annual cost is determined for each product $i = 1, \dots, m$ and distributor $j = 1, 2, 3$ according to the following relationship:

$$COST_{ij} = (1 + VAT)\left[12\left(mf_{ij} + mf_j + mo\right) + c\left(r^{HT}\left(p_{ij}^{HT} + p_j^{HT}\right) + r^{LT}\left(p_{ij}^{LT} + p_j^{LT}\right) + os + t\right)\right] \quad (1)$$

where

VAT … value added tax,

mf_{ij} … fix monthly fee of the product i and distributor j,

mf_j … fix monthly fee (for the circuit breaker) of the distributor j,

mo … monthly payment for other services,

c … annual consumption in MWh,

p_{ij}^{HT} … price in high tariff per 1 MWh of the product i and distributor j,

p_j^{HT} … price in high tariff per 1 MWh of the distributor j,

p_{ij}^{LT} … price in low tariff per 1 MWh of the product i and distributor j,

p_j^{LT} … price in low tariff per 1 MWh of the distributor j,

r^{HT} … percentage of the consumption in high tariff,

r^{LT} … percentage of the consumption in low tariff,

os … price for other services per 1 MWh (system services, support for electricity from supported energy sources)

t … electricity tax per 1 MWh (t = 28.3 CZK).

The final customer in the distribution area j should therefore choose the appropriate product i to minimize these annual costs:

$$I_j = \arg\min_i COST_{ij}, j = 1, 2, 3. \quad (2)$$

If we look at the problem from the perspective of multi-criteria decision making, it is evident that dominated products have no chance of success and a rational customer should never choose such a product. Unfortunately, the reality is often different and the household often does not know that its product is dominated, i.e. that there is a product with lower prices. However, it is sufficient for our model to perform its analysis only on a set of non-dominated products. Thus, for these products, the expected value of total annual costs according to (1) will be quantified and the most appropriate product selected according to (2).

In the case of known consumption c and known consumption ratios of low and high tariff, r^{LT} and r^{HT}, this is a simple decision making with certainty. But the reality is that we can only estimate annual consumption and ratio. For example, the simulation approach [8] can be used for the analysis, or the ratio (or the consumption) can also be used as a random variable to determine not only the best product at the expected consumption, but also the stability interval of this alternative. In other words, the aim of the following analysis will be to choose the most advantageous product and determine under which conditions the product will remain the best (cheapest) one.

Let us assume a fixed consumption c, e.g. 10 MWh and symbol r^{HT} denotes the ratio of the energy consumption in the high tariff to the total annual consumption. The determination of the interval of energy consumption ratio, for which the specific product k ($k = 1, \dots, 60$) is offered in the distribution network j ($j = 1,2,3$), is from a mathematical point of view gained by solving a system of optimization modes:

$$\begin{aligned} &\min (\max)\, r^{HT} \\ &\text{subject to} \\ &COST_{ij} \leq COST_{kj}, i = 1, \dots, 60. \end{aligned} \quad (3)$$

Note that the relation $r^{LT} = 1 - r^{HT}$ holds because the total annual consumption c is the sum of consumption in the low and high tariff.

If we solve the system (3) for the dominated product k, which is dominated by product i, it will be true that $COST_{ij} > COST_{kj}$ and the problem will have an infeasible solution. Thus, instead of 180 linear optimization problems (60 products for 3 distributors) with 60 linear constraints and two extremes (min and max) in every year, it is possible to solve only the problems for non-dominated products. In 2017, only 11 products were non-dominated for PRE and CEZ and 10 for E.ON and the optimization model (3) consists of 32 problems with 11 limiting constraints. In 2018, 13 products for PRE and ČEZ and 12 for E.ON were non-dominated and, according to (3), we solve 38 tasks with 13 limiting constraints.

4 Results

For the analysis, we assume the household with a fixed amount of annual energy consumption c about 10 MWh. The following values were substituted to the formula (1): $VAT = 0.21$, $t = 28.3$ and $c = 10$. As mentioned above, the relationship $r^{LT} = 1 - r^{HT}$ is also valid. In this case, the relationship (1) can be rewritten into form (4).

$$\begin{aligned} COST_{ij} = 14.52\left(mf_{ij} + mf_j + mo\right) \\ + 12.1\left(r^{HT}\left(p_{ij}^{HT} + p_j^{HT}\right) + (1 - r^{HT})\left(p_{ij}^{LT} + p_j^{LT}\right) + os + 28.3\right). \end{aligned} \quad (4)$$

For the year 2017 parameters $mo = 4.9$ and $os = 588.94$ were substituted into (4), and $mo = 5.4$ and $os = 588.63$ for the year 2018.

4.1 Results of the year 2017

Data set for the year 2017 consists of 60 products offered in the D25d distribution rate. However, only 11 from them are non-dominated products (for PRE and ČEZ), respectively 10 (for E.ON). So we solved 32 optimization problems with 11 constraints. For each product, we set an interval with the minimal and maximal ratio of consumption in the high tariff. The results were surprisingly the same in all distribution regions and intervals for ratios of the high tariff were identical. From Table 1, we can see that only four products can be the best in the case of 10 MWh consumption. For households with low consumption in the high tariff, the best product is Amper Market, HOME_AKU. It remains the best if the high tariff consumption is less than 26.76 % of the total annual amount of

electricity. In the case, the consumption in the high tariff is from 26.76 to 46.43 % one of CARBOUNION BOHEMIA product is the winner and the best choice for consumers. Households with higher consumption in high tariff should choose Nano Energies Trade s.r.o. product Dobrý skutek. The other products are for households with an average annual consumption of about 10 MWh inefficient.

From the previous analysis [10] we can determine, for example, that both CARBOUNION BOHEMIA products are the best not only in the case of consumption 10 MWh but also for all households with consumption higher than 10 MWh and the ratio $r^{HT} \in \langle 0.2676, 0.4643 \rangle$.

Ratio of high tariff [%]		The best product (PRE, E.ON, ČEZ) 2017
from	till	
0.0000	26.7606	Amper Market, HOME_AKU
26.7606	46.4341	CARBOUNION BOHEMIA, spol.s.r.o., STANDARD
		CARBOUNION BOHEMIA, spol.s.r.o., STANDARD 12
46.4341	1.0000	Nano Energies Trade s.r.o., Dobrý skutek

Table 1 Results of the optimization models for the year 2017

4.2 Results of the year 2018

Data set for the year 2018 consists of 13 non-dominated products (for PRE and ČEZ), respectively 12 (for E.ON). So, we solved 38 optimization problems (4) with 13 constraints. For each product, we also set the interval with the minimal and maximal ratio of consumption in the high tariff. The results (Table 2) are also the same in all distribution regions and intervals for ratios of the high tariff are identical again.

Ratio of high tariff [%]		The best product (PRE, E.ON, ČEZ) 2018
from	till	
0	34.0800	Europe Easy Energy, eDOMÁCNOST
34.0800	79.2877	Eneka s.r.o., Jednička
79.2877	1.0000	Eneka s.r.o., E-tarif

Table 2 Results of the optimization models for the year 2018

In 2018 only 3 products can be the best in the case of 10 MWh consumption (see Table 2). For households with the low high tariff consumption, the best product is Europe Easy Energy, eDOMÁCNOST. It remains the best if the high tariff consumption is less than 34.08% of the total annual amount of electricity. This border is higher than in the case of the year 2017. If the annual consumption in high tariff is from 34.08% to 79.29%, product Jednička from supplier Eneka s.r.o. is the best choice for the consumers. In other cases households should also choose Eneka s.r.o., but the best product is E-tariff. The other products were for households with an average annual consumption of about 10 MWh inefficient in 2018.

Based on linear optimization models, the intervals for the ratio of consumption electricity in high tariff and the corresponding cheapest products were obtained. According to these intervals, in 2017 there were 3 companies with 4 products. Also in 2018 three companies can be the best but the list of these companies is different. In both cases when the ratio for the high tariff consumption starts to be a little bit higher/lower than the limits of the intervals, it is not necessary to change the supplier and product in the given year as the increase in costs is not so big. However, our previous analysis [10] showed that one of the cheapest products for the year 2017 offered by the CARBOUNINON BOHEMIA company belonged between the most expensive ones in 2018. So we can conclude that the change in the energy supplier was suitable between 2017 and 2018. The question is: "Is the change significant?"

Let us assume 10 MWh consumption in PRE distribution and the ratio of high tariff consumption in Table 3. For the year 2017, the best tariff is displayed with total annual costs corresponding with the given product but also in the year 2018. In the next column, the minimal costs (for an optimal product in 2018) are displayed.

Nano Energies Trade s.r.o. in the year 2018 offers new product under the other part of the Nano Energies company called Nano Green, so the annual cost for 2018 was calculated for this product. As we have mentioned

above, the prices raised in the year 2018, and it is also evident from Table 3. When we compare the annual cost of 2017's best product in 2018 with the minimal costs product (two middle columns in Table 3), we see that the possible significant savings could be made by switching to a different product in 2018. When the consumption ratio in the high tariff rate is low (about 20%) the year-on-year increase in price (two right columns in Table 3) when the product is changed to the best 2018's is about 3.9% while the increase in cost without the product change is about 9.3% (by changing the product, the consumer saves 1 268.08 CZK). More pronounced is the change when 30% ratio is used – the annual cost can increase at about 37.6% when the product is not changed, but with the change, it can be only 4.5% increase (and the total costs are lower about 8 535.35 CZK). The situation is similar to other ratios.

Ratio	The best supplier 2017	Costs in 2017	Costs in 2018	Minimal costs in 2018	Cost increase	
					With change	Without change
20 %	Amper Market	23 577.77	25 764.92	24 496.84	3.9%	9.3%
30 %	CARBOUNION BOHEMIA	25 764.70	35 458.24	26 922.89	4.5%	37.6%
40 %	CARBOUNION BOHEMIA	27 835.47	37 627.78	29 259.42	5.1%	35.2%
50 %	Nano Energies Trade s.r.o.	29 794.92	36 336.66	31 534.23	5.8%	22.0%
70 %	Nano Energies Trade s.r.o.	33 312.10	41 267.41	36 083.86	8.3%	23.9%
85 %	Nano Energies Trade s.r.o.	35 949.98	44 965.47	39 243.79	9.2%	25.1%

Table 3 Results comparison

5 Conclusions

The number of products offered by electricity suppliers to households at the D25d distribution rate is very high (about 60 products offered by nearly 30 suppliers), but most are not suitable for any household as they are dominated products, i.e. products with higher prices than others. Nevertheless, many households still use them, even though it is possible to use the ERU calculator [5] and find a more suitable supplier, respectively product. Only about one-sixth of the products are among the non-dominated, i.e. they can be theoretically among the cheapest. The aim of our analysis was to find out for what ratio of the high and low tariff rates the given product was the cheapest one in the year 2017 and 2018 when the average annual electricity consumption was about 10 MWh. The results of the optimization models showed that the situation was changing every year and it was suitable to change the electricity supplier/product between the years as the cost savings could be about 9% or up to several thousand CZK. The analysis also showed that there were only a few products suitable for any ratio and any consumption level and the wide range of products offered was completely useless and rather confusing for customers.

Acknowledgements

The research project was supported by the grant No. F4/66/2019 of the Faculty of Informatics and Statistics, University of Economics, Prague and by the Czech Science Foundation (grant number 19-08985S).

References

[1] Boltz, W. (2013). The Challenges of Electricity Market Regulation in the European Union. *Evolution of Global Electricity Markets*. Elsevier Inc. 2013, 199-224.

[2] CEZ (2019). Energetika v ČR. [online], available: https://www.cez.cz/cs/pro-media/cisla-a-statistiky/energetika-v-cr.html [cit. 2019-04-18].

[3] Diakoulaki, D., Karangelis, F. (2007). Multi-criteria decision analysis and cost-benefit analysis of alternative scenarios for the power generation sector in Greece. *Renewable and Sustainable Energy Reviews*, *11/4*, 716–727.

[4] Elektrina.cz (2014). Distributor nebo dodavatel, kdo je kdo? [online], available at: https://www.elektrina.cz/distribuce-elektriny [cit. 2019-04-12].

[5] ERU (2019). Energy Regulatory Office - Price calculator, [online], available: http://kalkulator.eru.cz/VstupniUdaje.aspx [cit. 2019-04-12].

[6] Hegazy, Y.G., Salama, M.M.A. & Chickhani, A.Y. (2003). Adequacy assessment of distributed generation systems using Monte Carlo Simulation. *Power Systems - IEEE Transactions 18/1*, 48-52.
[7] Kuncová, M. (2015): Methods for the Electricity Supplier Selection – Case Study of the Czech Republic. *International Journal of Mathematical Models and Methods in Applied Sciences* **9** (2015), 714-720.
[8] Kuncová, M. & Sekničková, J. (2014): Analysis of the efficiency of the electricity supplier selection depending upon the price changes. *Proceedings of the 32th Mathematical Methods in Economics Conference* (pp. 542-547), Palacký University in Olomouc.
[9] Kuncová, M. & Sekničková, J. (2015): Optimization Models Used for the Electricity Supplier Selection, *Proceedings of the 33rd Mathematical Methods in Economics 2015 Conference* (pp. 443-448), West-Bohemia University, Cheb.
[10] Kuncová, M. & Sekničková, J. (2019). Selection of Electricity Supplier in 2018 in the Czech republic – Optimization Model. *Zborník zo seminára Využitie kv. metód vo ved.-výsk. činnosti a v praxi XIII*, Zuberec 2019 (pp. 98-106), Slovak Society for Operations Research, FHI EU, Bratislava.
[11] Newbery, D. (2013): Evolution of the British Electricity Market and the Role of Policy for the Low-Carbon Future. *Evolution of Global Electricity Markets*. Elsevier Inc. 2013, 3-30.
[12] OTE (2019). Basic information [online], available at: https://www.ote-cr.cz/cs/o-spolecnosti/zakladni-udaje [cit. 2019-04-10]
[13] Sekničková, J. & Kuncová, M. (2014). Evaluation of the locality influence on the electricity supplier selection. *Proceedings of the 32nd Mathematical Methods in Economics* (pp. 879-883), Palacký University in Olomouc.
[14] Ventosa, M., Baíllo, Á., Ramos, A. & Rivier, M. (2005): Electricity Markets modelling trends, *Energy Policy* **33**, 897–913.

Quantification of differentiating power in Network DEA

Michal Pieter[1]

Abstract. Network Data envelopment analysis (NDEA) allows one to analyze the internal structure of DMUs, thus better modeling the reality of the system. While the purpose of this is to better discriminate between the DMUs, the complexity of the model must also be taken into account. Common sense dictates that increasingly complicated models bring only diminishing returns after a certain point. And while the complexity of two models can be compared relatively easily, to determine which one is better suited, it is just as crucial to compare their power to differentiate between the DMUs. This paper thus looks at ways in which the differentiating power of NDEA models can be quantified for a given set of DMUs and then weighted against the complexity of those models.

Keywords: Network DEA, Data envelopment analysis, complexity, differentiating power, quantification.

JEL Classification: C38, C61, C65, C67
AMS Classification: 62H30, 90-06, 90-08, 90B10, 90C05, 90C60

1 Introduction

Data envelopment analysis (DEA) has since its inception by Charnes and Cooper [5] in 1978 (the CCR model) been regarded as a powerful and versatile tool for evaluating the performance of production units (decision-making unit, DMU) in a given group or set. It does this by computing their efficiencies with regards to inputs they consume and outputs they produce (together called factors), that are common to all the units. While many modifications had been developed since, most famously the BCC [4] and SBM [14] models, all of them treated the DMU as a black box, with inputs coming in and outputs going out, disregarding any potential internal operations. While an acceptable simplification in most cases, there are scenarios where these models are too far removed from reality and may give distorted results. Then, in 2000, Färe and Grosskopf [8] proposed an approach which allows one to model the internal machinery of DMU by introducing sub-processes, linked together by intermediary factors. This is the Network Data envelopment analysis (NDEA) approach and it has seen vigorous growth in the years since.

One potential issue in using NDEA instead of traditional, black-box models, is that there are often multiple ways in which the sub-processes for a given problem may be structured (more on this in chapter 1.1), even given the same set of factors. Moreover, the decision-maker may be tempted to bring the model as close to reality as possible, adding additional factors and constructing increasingly elaborate models with complicated structure. Obviously, this comes at a cost of additional complexity, making models more demanding to create, program and solve, more time-consuming to compute and more difficult to interpret. A balance then needs to be struck between this complexity and the quality of the results. Probably the most common criterion of this is the differentiating, or discriminating (terms can be used interchangeably) power of the model. [7] It is also the most sought after goal when coming up with new approaches to DEA. This stems from the issue that in most of the simplest models, many DMUs are assigned efficiencies that are the same or very similar. That is especially the case when the number of DMUs is small or the number of factors is large. The original CCR model is perhaps the most notorious offender in that case, often assigning most of the DMUs the efficiency of 1. Other models either do not fare much better, or are much too complex. That is why the goal of this paper is to try and find a way to balance the complexity and differentiating power between models, with special emphasis on Network DEA. Chapter 2 thus first starts with presenting various ways to measure complexity. In chapter 3 several ways to quantify the differentiating power of the model are proposed. Finally, a way to bring both measures together, with the goal of identifying a compromising model, is examined in chapter 4 and a corresponding measure is proposed. In that final chapter, some concluding remarks and a discussion of further research are also included.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq. 4, 130 67 Praha 3, Czech Republic, michal.pieter@vse.cz.

1.1 Network DEA structure

Arguably, the most important issue to consider when constructing a model in Network DEA is the relationship between the sub-processes and factors. Barring the trivial case of a single sub-process, which is equivalent to the black-box model (seen in figure 1), there are multiple ways of arranging them. In the simplest non-trivial case of two sub-processes, they can be arranged either in series or in parallel, as shown in figure 2. With more sub-processes, these two arrangements can be combined at will to produce more elaborate structures. Moreover, factors can also have more advanced relationships. They can be shared by more sub-processes, be undesirable, have saturation limits and so on. It is then up to decision-maker to choose one of these structures for their model.

Figure 1 Structure of traditional, black-box models

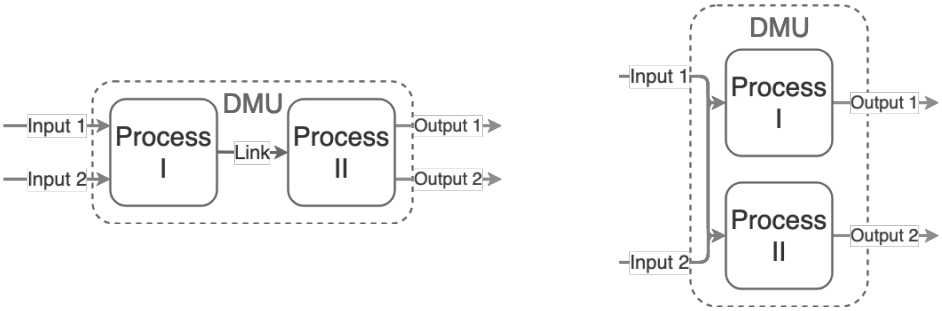


Figure 2 Example series and parallel network structures

While it may be tempting to think that the network structure should be objectively determined by the particulars of a given real system, or by economic theory, this need not always be the case. Consider the following case. In evaluating banks, or bank branches, 8 factors were identified and their values obtained for each bank. A 2-stage network model was constructed, as seen in figure 3. This structure fits common sense, as well as business practice. On closer inspection though, another sub-process can be inserted in the middle, making the model more precise, but also more complicated, as seen in figure 4. One can even argue that Equity should serve as input to all three sub-processes (larger banks can afford better marketing, people may equate size with stability when saving or borrowing, larger banks can hire better investors and enforce timely fee payment), making for an even more complex, mixed series-parallel structure, but one that is more likely to reflect the actual underlying reality. There is almost always a way to create multiple models, where each stands as a reasonable representation of the real system, with its own drawbacks and advantages. Clearly, the issue then arises as to how to choose between them.

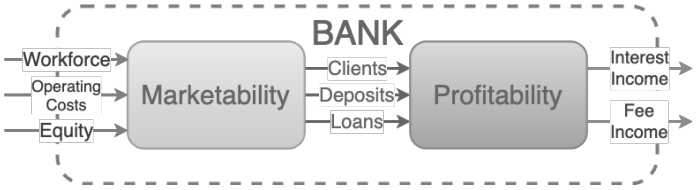


Figure 3 Network structure with 2 stages connected in series

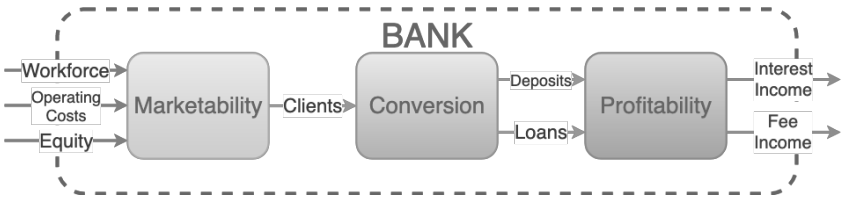


Figure 4 Network structure with the same factors but an additional middle stage

2 Quantifying complexity

Surprisingly little has been written about quantifying the complexity of not only Network DEA models, but also linear programming problems in general, with what little research in this area done being mainly focused on mixed-integer linear programming (MILP). It seems to be one of those "know it when you see it" issues, where in practically each case in the reviewed literature, the authors of a new model state that theirs is more complex, when

compared to some original, unmodified version. These are invariably qualitative statements and are self-evident when considering how these new models came to be – by adding more stages, conditions, variables, etc. Still, there are several ways of quantitatively measuring, crude as some of them are, the complexity of linear programming models in general and of Network DEA models in particular, with some of them outlined below. A disclaimer is needed though - this list is by no means exhaustive, more advanced quantifiers of complexity can be constructed, but is sufficient enough so that some of the proposed metrics can be used in the compromising measure to follow.

2.1 Number of variables and/or constraints

A consequence of the simplex method algorithm and its various more advanced modifications, is that the more variables are there in a model (linear and non-linear), and the more constraints must be satisfied – all other things being equal – the more difficult it is to obtain the optimal solution. MILP and other non-linear problems are especially sensitive to number of variables when solved using the Branch and Bound algorithm. And while in some cases additional constraints can reduce complexity, e.g. various relaxation and bounding approaches, the number of both variables n and constraints m can still serve as a crude approximation to model complexity. Because most NDEA models are linearized before computation, it might therefore be useful to consider this number, let's say denoted by $\Gamma_{VC}=n+m$, when comparing different models.

2.2 Number of factors

Another possible measure of complexity may be number of factors (inputs, outputs and intermediaries). While the addition of more data, in the form of factors, may make the model more precise, it also results in new constraints needing to be satisfied. As just discussed, this results in greater complexity. While it is ultimately just a proxy to Γ_{VC}, it is more readily obtainable, just by examining the structure of the model. However, it can also be misleading, when two models have the same number of factors but one has them for example spread out over more sub-processes, as is the case with figure 3 and figure 4. One must also consider shared factors – their effect is directly proportionate to the number of sub-processes linked, e.g. in the parallel model in figure 2 one must count each input twice. Counting up in this manner all inputs, outputs and intermediate factors, let us denote that number Γ_F.

2.3 Number of sub-processes

Similar to how number of factors can serve as a proxy for number of variables and constraints, so can the number of sub-processes. Some models, such as [9] compute efficiencies for each sub-process as a separate program, so that each new sub-process means another run. Other models, such as in [11] compute the overall efficiency and sub-process efficiencies at the same time. In that case, adding new sub-process means adding new constraints to the model. Another measure could therefore be the number of sub-processes, denoted by Γ_P.

2.4 Number of stages and degree of parallelism

While the number of sub-processes does determine the complexity of the model, just as important is their arrangement. The simplest cases are when they are all in series or parallel. When both types of structure are present – so called hybrid structure – it becomes harder because at points of divergence/convergence ratio constraints for shared factors must be added, as seen in for example [8] or [1]. To reflect this, one possibility is to identify all the series in the network and count up their stages (even if a stage contains multiple sub-processes in parallel), denoted by $\Gamma_D{}^S$. Then, all the parallel sub-structures are identified, the number of levels in each of them (even if it contains a series of sub-processes) is counted and added together, denoted by $\Gamma_D{}^P$. Let us then denote the total degree of serialization and parallelism as $\Gamma_D=\Gamma_D{}^S+\Gamma_D{}^P$.

2.5 Computational time

The most precise measure of complexity of a model is the actual time it takes to calculate a solution, let's say denoted by Γ_T. While it is notoriously difficult to estimate the computational time of a given model, there is another option. For model choice, several models can be run on only partial data to determine their Γ_T, and evaluated by one of the techniques outlined further in this paper. The one that performs the best is then selected and used to obtain results on the entire dataset. This measure of course is not perfect. It needs the actual data and so is dependent on it, as well as on the chosen software, where different solvers are optimized for different kinds of problems. It also does not reflect the underlying theoretical complexity and cannot distinguish for example between a simple yet inefficient model (perhaps non-linear) from a more complex but also more efficient. It is clear though, that a positive correlation between complexity and computational difficulty exists. Furthermore, it may be the most practical measure of all, especially in case of larger data sets, when computational time exceeds that of program preparation, and when multiple runs are expected.

3 Quantifying differentiating power

While there are only a few articles that attempt to quantify the complexity of DEA models, when it comes to their differentiating power, the situation as of writing this work is even worse. Try as one might, there appears to have been no research into this matter whatsoever thus far. While there are a great many cases where the authors ascribe superior differentiating power to their newly proposed models, it is invariably treated as an obvious result, unwarranted for further discussion. In the surveyed literature of various NDEA models, e.g. in [11], this comparison is always done through counting the number of efficient DMUs (as seen below), when applied to some practical problem. This number is then compared to the same metric given by a simple, non-network DEA model, usually CCR or BCC. Following are some proposed methods that attempt to actually quantify this quality in a way where it can be useful to compare arbitrary NDEA (or as a special case, simple DEA) models.

All of the proposed methods work by analyzing the computed efficiency scores. It is thus essential to first apply the model to an actual problem and calculate the results. This may either be a dummy problem when comparing the models in purely theoretical setting, or a practical problem with real data. In the latter case, a model with the best differentiating power may be then used and the results interpreted. When it is desirable to choose a model before the calculation happens, e.g. when it is expected to take a long time, it may be prudent to only take a sample of the data for model choice, and then calculate with full dataset, once the best model is identified.

3.1 Number of efficient DMUs

Is is a well known feature of traditional DEA models, such as CCR and BCC, but also Tone's SBM model, that they assign full efficiency scores to more and more DMUs, as the size of the problem increases. [7] This may be due to greater number of units compared, or due to additional factors taken into account. This number also increases when one allows for variable returns to scale, changing the shape of the production possibility frontier from conical to generally convex, as seen for example in BCC model. When the goal is to better discriminate between DMUs, this is an undesirable feature, hence why many of the earliest modifications sought to produce a full ranking of units, reducing the number of those efficient to 1, for example the Andersen-Petersen super-efficiency model[2]. For ease of comparison, let us then define $\Delta_N = 1 - n_E/n$, where n is the total number of DMUs and n_E the number of efficient ones. Greater values of Δ_N therefore imply greater discriminating power. A slight modification may allow cases where a DMU is so close to being efficient, that any reasonable decision-maker would consider it as such. If we define α as a tolerance factor, let's say α=5%, then $N_E{}^\alpha$ is the number of DMUs that come at least α close to full (1) efficiency, with respect to the efficiency range of all units. $\Delta_N{}^\alpha$ can therefore be defined as

$$\Delta_N^\alpha = 1 - {}^{n_E^\alpha}/_{n}, \qquad n_E^\alpha = |E_\alpha|, \qquad E_\alpha = \left\{ i, i \in \mathrm{N}: \frac{1 - E_i}{1 - \min_{j \in N}(E_j)} \leq \alpha \right\},$$

where N is the set of all DMUs, E_k is the computed efficiency score of k-th DMU and E_α is the set of tolerance-adjusted efficient DMUs. There are models that can assign above-unity scores, such as aforementioned AP[2] or Tone's SSBM[15], so when these need to be evaluated, the unity score is simply replaced with $\max_{j \in N}(E_j)$.

3.2 Variance

Another possible measure is the spread of computed efficiency scores, as measured by either their variance or standard deviation. Say $\Delta_S = \sqrt{\sum(E_k - \mu)^2/n}$, where μ is the mean efficiency score. The expectation is that the less discriminating power the model has, the more the efficiency scores will be concentrated around a single value, in all likelihood 1, and the harder it will be to distinguish between DMUs. Conversely, if computed efficiency scores are more evenly spread out, there will be greater confidence with which DMUs can be distinguished. Ideally, only those DMUs that are truly similar in their efficiency would get similar scores, and the rest would be uniformly spread between 0 and 1 (if one does not consider above-unity scores). There is a caveat though. There can be relatively few values, very different from each other, around which all efficiency scores would be concentrated. Consider the extreme case where several DMUs would get (almost) full efficiency score and the rest would get (almost) zero efficiency score. While standard deviation would be as high as it could be, and yes, the two groups would be clearly distinguishable, within each group the discriminatory power would be almost none.

3.3 Cluster analysis

To safeguard against previously described scenario, one of the tools of cluster analysis could be applied on the computed efficiency scores. There are many algorithms that try to identify clusters of data points, which share similar values of one or more variables. It is the ability to analyze data in multiple dimensions that make these especially useful when dealing with Network DEA models. Examples include the popular k-means algorithm [10],

Gaussian mixture models [13] and others [12]. Univariate algorithms include the Jenks natural breaks optimization, which has the benefit of providing globally optimal clustering. In that case, the only dimension of data is obviously the overall network efficiency score. With multivariate methods, one can include at will more dimensions that represent sub-process efficiencies. Then, one of the techniques of cluster analysis can be used to identify the best fitting clusters, with the precise meaning of "best" depending on the chosen algorithm.

There are two ways of identifying the clusters. Most algorithms require the number of clusters as an external input parameter. If not given, some try to determine this number through a heuristic, such as the "elbow" criterion for k-means, or Yan's Gap method [16]. Alternatively, if the number of DMUs n is not very large, e.g. when using only sample data, the algorithm can be run exhaustively for all possible numbers of clusters, from 1 to n. Either way, a number of identified clusters, say k, can be determined this way, providing a measure of sorts. Clearly, if fewer and better-fitting clusters are identified, that means the DMUs have similar efficiency scores, and are therefore less differentiable. Conversely, a great number of clusters (ideally, $k=n$), or a few but poorly-fitted clusters, mean that the DMUs are not similar at all and are easily distinguishable. To quantify this degree of clustering, a measure proposed here, let's call it Δ_C, can be defined as follows:

$$\Delta_C = \frac{1}{k}\sum_{j=1}^{k}\frac{d_j+\varepsilon}{d_j^{\times}+\varepsilon}, \qquad \Delta_C \in (0;1\rangle,$$

where d_j is a degree of clustering of cluster j, $d_j^{\times}$ is the worst possible degree of clustering cluster j could have, were its data points not at all clustered, and ε is a non-Archimedean infinitesimal. Obviously, d_j and $d_j^{\times}$ depend on the chosen algorithm, for example if k-means were used, d_j would be actual intracluster variance and $d_j^{\times}$ would be this variance, were all the data points the maximum distance from the cluster centroid. There are other possible measures, of course. One could for example use the technique of Anderson [3], specifically his p-value, derived through multiple permutation-based quasi-Fisher tests, which compares the "distinguishability" of previously identified clusters. And while it is more robust, it requires many calculations – $(kn)!/k!(n!)^k$, or a sufficiently large sample thereof – so is generally only suited for relatively small number of DMUs.

While cluster analysis offers a robust way to quantify the differentiating power, especially with multivariate algorithms, it has its drawbacks. While many of them are included in popular statistical programs, such as R, SPSS, Stata and others, the computational difficulty increases precisely with the number of dimensions, i.e. sub-processes. Furthermore, the majority of these algorithms guarantee at best only local optima.

3.4 Entropy

There are other approaches that can be successfully adapted to quantify the differentiating power of NDEA models. For example, Chen et al. [6] proposed a smoothed entropy-based measure of multivariate spread. While their analytical formulae require the knowledge of probability density functions, they go on to provide estimators for finite samples as well, based on kernel density estimators. They provide two measures, Joint Entropy (JE) and Total Marginal Entropy (TME) that converge in value with increasing smoothing factor λ, as follows:

$$\widehat{H}(E) \approx \hat{T}(E) \approx P'\log\left(\lambda\sqrt{2\pi}\right) + \frac{1}{\lambda^2}\sum_{p=1}^{P'}\hat{\sigma}^2_{E^{(p)}}, \qquad \hat{\sigma}^2_{E^{(p)}} = \frac{1}{n}\sum_{i=1}^{n}\left[E_i^{(p)} - \frac{1}{n}\sum_{j=1}^{n}E_j^{(p)}\right]^2,$$

where $\widehat{H}(E)$ is JE, $\hat{T}(E)$ is TME, while P' is the number of samples (variables, dimensions) and $\hat{\sigma}^2_{E^{(p)}}$ the estimated variance of observations from sample p, or in the case of NDEA, of efficiencies of p-th sub-process. P' is the dimensionality, so in the case of NDEA either the number of sub-processes P, or $P+1$, when one chooses to include also the total efficiency score. Chen et al. argue that for $\lambda \geq 2$, a sufficient convergence of JE and TME is achieved. Therefore, setting $\lambda=2$, a single measure of entropy, called in the context of NDEA say Δ_E, can be obtained. When calculated for two or more models, these can be compared. Since low entropy indicates less information needed to describe the system, i.e. distinguish the states (efficiency scores), the models with larger relative values of Δ_E can be interpreted as having greater differentiating power.

4 Compromising measure and concluding remarks

In previous chapters, several measures have been proposed, both for the complexity of the model, as well as its differentiating power. For model selection and evaluation, a compromise between these two attributes must be reached, therefore it seems desirable to create a combined metric. Such a measure would be able to determine if a model, that is say more complex (as given by a chosen measure) could still be more advantageous to use, provided it would result in achieving disproportionately greater differentiating power. Suppose there are R models being investigated, each having been computed on a subset of the data. One with the greatest differentiating power, with regards to its complexity, is to be chosen for the full calculation. Now suppose that $\Gamma^{(r)}$ is the estimated complexity

of r-th model, using one of the proposed techniques, and $\Delta^{(r)}$ is the likewise-obtained differentiating power. The most obvious combined measure is the ratio $\Delta^{(r)}/\Gamma^{(r)}$ – note however, that none of the measures are on the same scale (with the exceptions of Δ_N and Δ_C). Instead, their relative values may be taken, with regards to all models. Furthermore, consider the case where thus obtained ratio is the same (or close to) for two models, but with one having low complexity and differentiating power, and the other both of them high. Seemingly equivalent, in real world the objective factors or subjective preference will likely favor either decreased complexity or increased differentiating power. By introducing a weighting parameter $\alpha \in (0;1)$, this preference can be taken into account. The greater the α, the stronger preference there is for differentiating power, with $\alpha = 0.5$ meaning no preference either way. Proposed compromising measures of Π' (no preferences taken into account) and Π'_α (with preferences), where Π' is defined as $\Pi' = \Pi'_{0.5}$, can then be constructed as below. Note that $\Pi'_\alpha \in (0;1)$, with higher values implying greater relative differentiating power with respect to estimated model complexity.

$$\Delta'^{(r)} = \Delta^{(r)} \Big/ \sum_{s=1}^{R} \Delta^{(s)}, \qquad \Gamma'^{(r)} = \Gamma^{(r)} \Big/ \sum_{s=1}^{R} \Gamma^{(s)}, \qquad \Pi_\alpha^{(r)} = \frac{\Delta'^{(r)\ w_\alpha}}{\Gamma'^{(r)\ 1/w_\alpha}}, \quad w_\alpha = \frac{\alpha}{1-\alpha}, \qquad \Pi'^{(r)}_\alpha = \frac{\Pi_\alpha^{(r)}}{\sum_{s=1}^{R} \Pi_\alpha^{(s)}}.$$

The advantages of measures proposed in this paper lie in quantifying, to at least some level, both the complexity and differentiating power, and then combining them into a single measure, allowing one to evaluate and choose suitable model from several in consideration. Which particular measures are chosen affects the accuracy of such quantification, as well as difficulty of obtaining it. The disadvantage, when it comes to measures of differentiating power, is their dependence on actual data, requiring calculation of the model, if only on a sample. It remains an open question, whether a generic measure could be devised, dependent only on some fundamentals of the model. Especially considering the lack of literature on the subject discussed here, more research is needed. It would also be desirable to see a practical application of some of the proposed measures to a real world model-choice scenario. Regrettably, such an exercise could not be included in this paper at the time of publication. This issue is to be rectified in a follow-up paper, which is being prepared and is planned to be published soon.

Acknowledgements

This work was supported by grant no. IGA F4/20/2018 of the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Amirteimoori, A., and Yang, F.: A DEA model for two-stage parallel-series production processes. *RAIRO - Operations Research* **48** (2014), 123–134.
[2] Andersen, P., and Petersen, N.C.: A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science* **39** (1993), 1261–1264.
[3] Anderson, M.J.: A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26** (2001), 32–46.
[4] 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.
[5] Charnes, A., Cooper, W.W., and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[6] Chen, B., Wang, J., Zhao, H., and Principe, J.C.: Insights into Entropy as a Measure of Multivariate Variability. *Entropy* **18** (2016), 196.
[7] Despotis, D.: Improving the discriminating power of DEA: Focus on globally efficient units. *Journal of the Operational Research Society* **53** (2002), 314–323.
[8] Färe, R., and Grosskopf, S.: Network DEA. *Socio-Economic Planning Sciences* **34** (2000), 35–49.
[9] Kao, H.-Y., Chan, C.-Y., and Wu, D.-J.: A multi-objective programming method for solving network DEA. *Applied Soft Computing* **24** (2014), 406–413.
[10] Lloyd, S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28** (1982), 129–137.
[11] Lozano, S., Gutiérreza, E., and Salmerónb, J.L.: Network DEA models in transportation . Application to airports. 2009.
[12] Madhulatha, T.S.: An overview on clustering methods. *IOSR Journal of Engineering* **02** (n.d.), 719–725.
[13] Reynolds, D.: Gaussian Mixture Models. In: *Encyclopedia of Biometrics*, (Li, S.Z., and Jain, A., eds.), *Springer US*, Boston, MA, 2009, 659–663.
[14] Tone, K.: A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* **130** (2001), 498–509.
[15] Tone, K.: A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research* **143** (2002), 32–41.
[16] Yan, M.: Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion. (2005).

Development of Parliamentary Voting of Deputies without Party Affiliation in the Czech Parliament

Elena Mielcová[1]

Abstract. In the Czech Parliament, during the parliamentary period, there are usually several changes in party affiliation of parliamentary deputies. Some deputies change political party, some of them become deputies without party affiliation. The deputies without party affiliation can vote with respect to their old party, or with respect to any other political party. Their political views can develop over time, sometimes after time these deputies without party affiliation choose to join another political party. The main aim of this article is to study a closeness of selected deputies in the Chamber of Deputies of the Parliament of the Czech Republic to the voting of deputies of all political parties in session-by-session comparison during the selected time-period.

Keywords: Czech parliament, party affiliation, votes.

JEL Classification: D71
AMS Classification: 62H30

1 Introduction

In the Chamber of Deputies of the Parliament of the Czech Republic, all 200 deputies are elected with respect to their party affiliation or association with the political party. During the parliamentary period, deputies of the Chamber of Deputies are members of political clubs with respect to their political party affiliation or association:

"*The legal delimitation of the activities of the political groups can be found in the Act on the Rules of Procedure of the Chamber of Deputies. Deputies can associate in political groups, namely according to membership in the political parties or political movements for which they were candidates in the elections. Deputies who are members of one political party can form only one political group.*" (cited from [4]).

During the parliamentary period, there are usually several changes in political club affiliations; these changes are usually, but not necessarily, connected with change in political party affiliation. Deputies can join another political club or can become independent deputies. The process of change is not easy and the incentive to change political party affiliation should be strong. When this behavior of deputies occurs, some questions could arise: Do these deputies have different political views (expressed by their voting) comparing to other political party members before they decide to leave the political party? Do their votes evolve over time? Hence, this is the motivation behind the main aim of this article, which is to study the development of voting of selected legislators, who were without political party club affiliation during the selected parliamentary period.

The selection of the appropriate parliamentary period was quite easy; the most interesting and politically challenging situation occurred after 2006 parliamentary elections, when both potentially political coalition and opposition parties gained 100 deputies. Thus, the elections produced a deadlock situation with less than the 101 votes in both sides required to pass legislation or to pass a confidence vote for the new government in the Chamber of Deputies [3]. Finally, on January 19, 2007, a coalition government passed a confidence vote, when two deputies, originally elected as Social Democrats (CSSD) abstained. These two deputies had left their political clubs well before this vote; their votes in session-by-session comparison before and after the confidence vote are a topic of this study.

This article is organized as follows: the description of data used is given in the next section. Section 3 covers methodology of vote classification; in this case, the method of k-nearest neighbors was used. Results of calculations are presented in Section 4; this part also explains the choice of estimation parameters. Conclusions with the list of references ends the article.

[1] Silesian University in Opava / School of Business Administration in Karvina, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech republic, mielcova@opf.slu.cz.

2 Data Description

As mentioned in previous text, the data composes of votes of all legislators in the Chamber of Deputies of the Parliament of the Czech Republic from the period after 2006 elections. In general, after 2006 elections, there were five political parties operated in the Chambers of Deputies of the Czech Parliament. Three political parties formed governmental coalition:

- Civic Democratic Party (ODS) – 81 seats after the elections;
- Christian and Democratic Union - Czechoslovak People's Party (KDU-CSL) – 13 seats after the elections;
- Green Party (SZ) – 6 seats after the elections.

Two remaining political parties stayed in the opposition:

- Czech Social Democratic Party (CSSD) – 74 seats after the elections;
- Communist Party of Bohemia and Moravia (KSCM) – 26 seats after the elections.

Basic information on the Czech Parliamentary system as well as the set of all historical votes can be found at the official web site of the Chamber of Deputies of the Parliament of the Czech Republic [5].

As mentioned before, there were two deputies of CSSD, who had left their political party club before the confidence vote in January 2007. For simplicity, the text refers of them as deputy A and Deputy B. Deputy A had left party parliamentary club before December 8, 2006, in the middle of Session 7 (compare votes [8] and [9]). Deputy B had left party parliamentary club before October 25, 2006, in the middle of Session 6 (compare votes [6] and [7]). Hence, in session-by-session comparison Sessions 6, and 7 were split into two parts. For this study, all calculations are performed on data from sessions 6 – 14 with exceptions of sessions when deputies A and B were not present. Session 12 composed of only 3 votes, and took place in the middle of session 11, thus these three votes are covered in results of Session 11. Overview over data used is given in Table 1.

Sessions	Dates of Sessions	Number of votes	Legislators A and B present
Session 6 – part 1	24/10/2006 – 24/10/2006	31	yes
Session 6 – part 2	25/10/2006 – 8/11/2006	183	yes
Session 7 – part 1	28/11/2006 – 6/12/2006	125	yes
Session 7 – part 2	8/12/2006 – 13/12/2006	457	yes
Session 8	8/12/2006	11	no
Session 9	19/1/2007	4	no
Session 10	17/1/2007	2	no
Session 11	30/1/2007 – 8/2/2007	192	yes
Session12	7/2/2007	3	yes
Session 13	13/3/2007 – 21/3/2007	146	yes
Session 14	24/4/2004 – 10/5/2007	301	yes

Table 1 Overview over data used in calculations. Deputies A and B were not present during sessions 8, 9, and 10. Session 12 composed of only 3 votes, and took place in the middle of session 11, thus these three votes are covered in results of Session 11 votes. Source: own overview based on information from the official web site of the Chamber of Deputies of the Parliament of the Czech Republic [5].

All data cover votes of deputies on various topics. The result of every voting could be "yes", "no", "present, abstain", "absent", and "absent, excused". For the purposes of calculations all results were re-evaluated such that:

- "yes"=1,
- "no"=0,
- "present, abstain"=0,
- "absent"=0.5, and
- "absent, excused"=0.5.

In this analysis the "present, abstain" outcome is classified the same as "no" outcome. Even though some authors (e.g. Freixas, Zwicker [1]) propose to take into account three expressed legislators' outcomes (that means

"no", "yes", and "present, abstain"), the "present, abstain" outcome in voting in the Czech Parliament is influencing result of voting the same way as "no" outcome. The re-evaluations were done in a way that no normalization for purposes of k-nearest neighbors' analysis is needed.

3 Methodology

As mentioned above, the classifications of political views of selected deputies A and B is given by k-nearest neighbors' analysis. In general, the k-nearest neighbors' analysis is a method for classifying cases based on their similarity to other cases [2]. Classification determines the class label for an unlabeled test case. Cases that are near each other are said to be "neighbors". The unknown cases are labelled with respect to their k neighbors, which are the closest ones. The general algorithm:

1. Pick a value for k.
2. Calculate the distance of unknown case from all cases.
3. Select the k observations in the data set that are "nearest" to the unknown data point.
4. Predict the classification of the unknown case using the most frequent label value from the k nearest neighbors.

This way we can determine which political parties (=labels) are closest to classified deputies. Results are dependent also on selected distance calculations; in this analysis the Euclidean distance was used. The determination of the best k can be done by comparison of so-called Jaccard index for more values of k.

In general, the evaluation metrics in classification is based on splitting the set into the train set and the test set. Then the actual label values in the test set can be compared with the values predicted by the model in order to calculate the accuracy of the model. Jaccard index (called also Jaccard similarity coefficient) compares similarities of actual labels y with predicted labels $\hat{y}$:

$$J(y,\hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} = \frac{|y \cap \hat{y}|}{|y| + |\hat{y}| - |y \cap \hat{y}|} \quad (1)$$

where $|y|$ is a cardinality of y. Values of Jaccard index can be from interval [0,1]. When $J(y,\hat{y}) = 0$ then there is no match between actual labels and predicted labels. In the case when $J(y,\hat{y}) = 1$, then all predicted labels are predicted correctly.

All calculations were done using scikit-learn, a free machine learning library for the Python programming language. Basic calculations were done for the set number of neighbors $k = 10$; the determination of the best k were based on calculations for values $k = 1, k = 2, \dots, k = 14$. Use of higher values of k is counterintuitive because the smallest political party (SZ) received only 6 seats in the parliament and thus cannot be the label in the case when $k \geq 12$.

4 Results of Calculations

As mentioned in previous session, deputies A and B influenced the vote on confidence of government by their absence in the meeting, namely their absence during Session 9 held on January 19, 2007. The classification of their votes cover approximately three months before and four months after this vote.

With respect to the votes, the k-nearest neighbors classification of their voting pattern was performed for each session; in this case number of nearest neighbors was set to be $k = 10$. In order to evaluate the accuracy of classification, the data set containing votes was split into two parts; 80% of data were set to be the train set, the rest was expected to be a test set in order to estimate the out-of-sample accuracy of the analysis. This split was determined such that 40 deputies with known labels (party affiliation) were let for out-of-sample classification and comparison. Results of classifications as well as values of train-set accuracy and test-set accuracy calculated as Jaccard index are given in Table 2.

Calculations show that both deputies A and B voted with respect to their original party affiliation during sessions preceding confidence vote (Sessions 6 and 7). However, their votes started to change after the vote; interestingly their votes are closest to the CSSD long-term opponent political party – ODS.

Session	Classification of Deputy A	Classification of Deputy B	Train Set Accuracy	Test Set Accuracy
Session 6 – part 1	CSSD	CSSD	0.800	0.725
Session 6 – part 2	CSSD	CSSD	0.931	0.875
Session 7 – part 1	CSSD	CSSD	0.925	0.800
Session 7 – part 2	CSSD	CSSD	0.819	0.675
Session 11	CSSD	CSSD	0.919	0.875
Session 13	ODS	CSSD	0.906	0.825
Session 14	ODS	ODS	0.906	0.825

Table 2 Classifications of votes of deputies A and B with respect to k-nearest neighbors method for given value of neighboring observations $k = 10$. Last two columns contain train and test accuracies calculated as Jaccard indexes. Source: own calculations.

In presented results, the number of nearest neighboring observations used in classification was given as a fixed number, in this case $k = 10$. This number was chosen because it is big enough to block outliers and small enough to cover the political party with smallest number of seats – in this case SZ (Green Party) with six seats in the Chamber of Deputies.

Might happen that this result would change, if the number of neighbors is set to be the best value of k. To test this opportunity, the calculation of out-of-sample accuracy was calculated for different values of k, namely values from 1 to 14. For every k, the test set accuracy and its standard deviation was calculated. Value k with highest accuracy was chosen to classify deputies. For illustration, Figure 1 gives values of test set Jaccard indexes for all calculated values of k for the case of Session 6. In this case, for the first part of Session 6 the optimal number of nearest neighbors is $k = 6$. For this value of k both deputies are classified as voting with respect to CSSD.

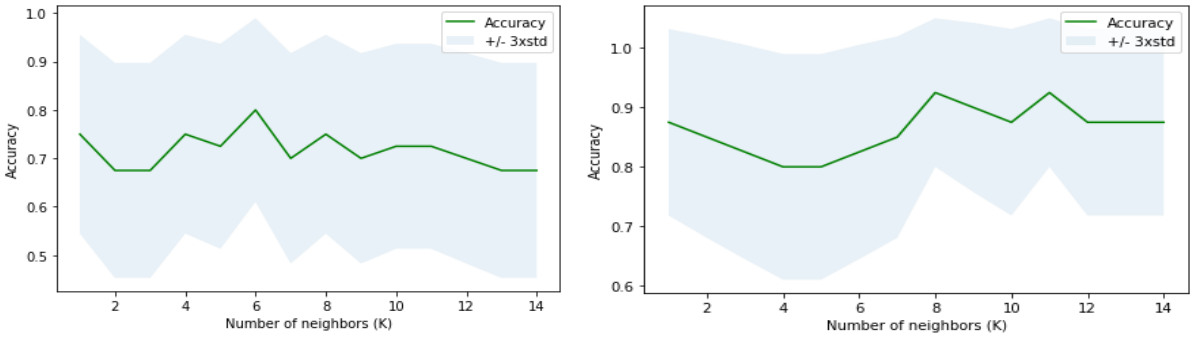


Figure 1 Values of Jaccard indexes for values of $k = 1,2, \ldots ,14$ for the first part (left) and the second part (right) of Session 6. Source: Own calculations.

For the second part of Session 6 the optimal number of neighbors is $k = 8$ or $k = 11$ (see also Figure 1), this part of session cover more than six times bigger data set comparing to the first part of Session 6. However, for both values of k both deputies are classified as voting with respect to CSSD.

The optimal number of neighbors for the Session 7 is $k = 4$ or $k = 12$ for the first part of data; for both values of k both deputies are classified as voting with respect to CSSD. For the second part of Session 7, the optimal number of neighbors is $k = 1$ and $k = 14$ (see Figure 2). In the case of $k = 14$ both deputies are classified as voting with respect to CSSD. The possibility $k = 1$ nearest neighbor observation was excluded from this analysis..

The interesting situation occurred when, in the case of Session 11. The best result is received in the case when $k = 1$ and accuracy is equal to 0.9, however this increases the risk of outlier observations. The second best option reveals that multiple possible choices are optimal for selections of k (possibilities $k = 3$, $k = 6$, $k = 8$, and $k = 10$ has the same out-of sample accuracy $J(y, \hat{y}) = 0.875$ (see figure 3). For purposes of classification, only the possibilities $k = 3$, $k = 6$, $k = 8$, and $k = 10$ were considered, under this condition both deputies are classified as voting with respect to CSSD.

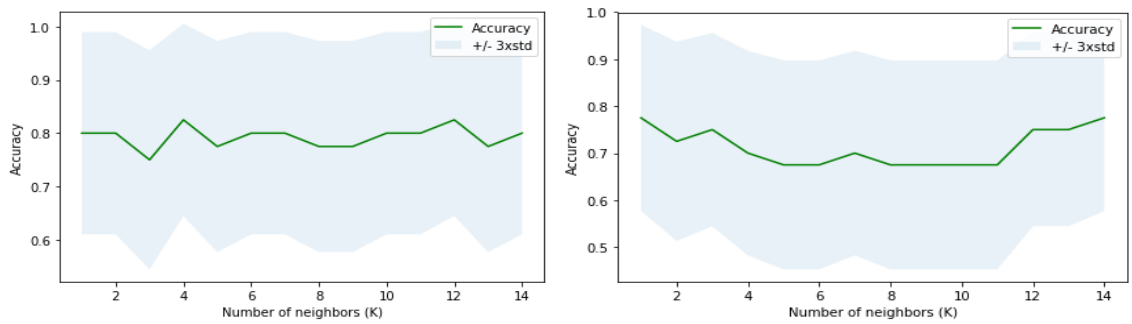


Figure 2 Values of Jaccard indexes for values of $k = 1,2, \ldots ,14$ for the first part (left) and second part (right) of Session 7. Source: Own calculations.

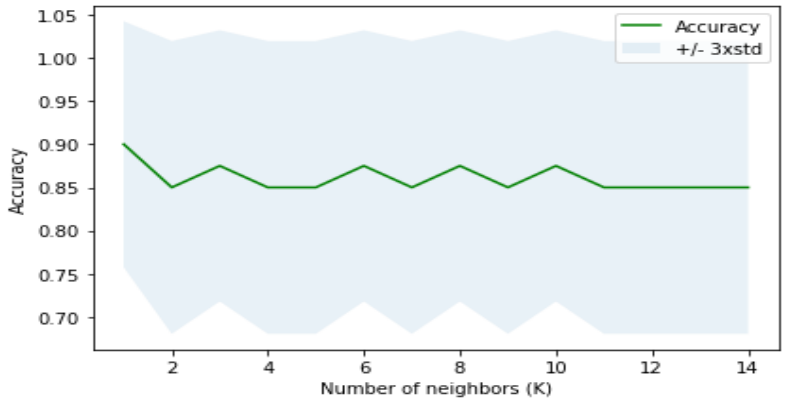


Figure 3 Values of Jaccard indexes for values of $k = 1,2, \ldots ,14$ in the case of Session 11 data set. This data set cover also observations from Session 12. Source: Own calculations.

In the case of data from Session 13 voting, the ideal value of neighbors is given as $k = 1$ or $k = 3$; in both cases the accuracy index is $J(y, \hat{y}) = 0.85$ (see Figure 4). In the case when $k = 3$, deputy A is classified as KDU-CSL while deputy B is classified as voting with respect to ODS.

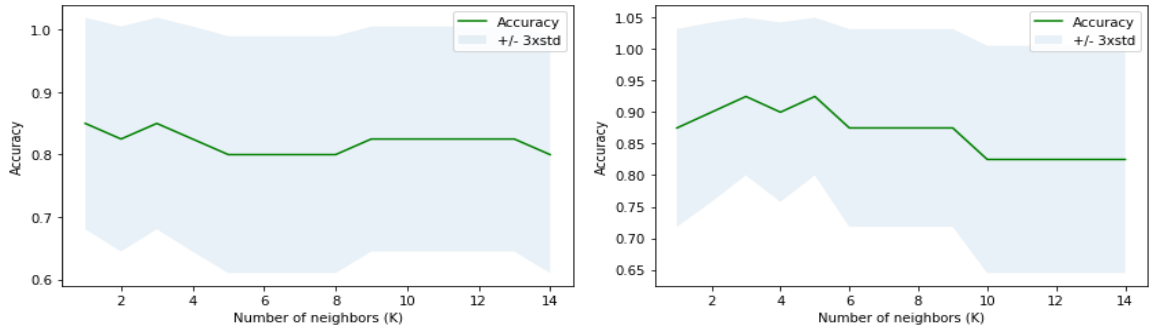


Figure 4 Values of Jaccard indexes for values of $k = 1,2, \ldots ,14$ in the case of Session 13 (left) and 14 (right) data sets. Source: Own calculations.

The last studied voting data set from Session 14. In this case the ideal number of neighbors is $k = 3$ or $k = 5$; in both cases the accuracy index is $J(y, \hat{y}) = 0.925$ (see Figure 4). For both values of k both deputies are classified as voting with respect to ODS. Results of classifications of deputies A and B are summed up in Table 3.

Session	Classification of Deputy A	Classification of Deputy B	Optimal k	Test Set Accuracy
Session 6 – part 1	CSSD	CSSD	6	0.8
Session 6 – part 2	CSSD	CSSD	8, 11	0.925
Session 7 – part 1	CSSD	CSSD	4, 12	0.825
Session 7 – part 2	CSSD	CSSD	14	0.775
Session 11	CSSD	CSSD	3, 6, 8, 10	0.875
Session 13	KDU-CSL	ODS	3	0.85
Session 14	ODS	ODS	3, 5	0.925

Table 3 Classifications of votes of deputies A and B with respect to k-nearest neighbors method for optimal value of neighboring observations k. Last column contains test set accuracies calculated as Jaccard indexes. Source: own calculations

Results support observations from Table 2: both deputies A and B voted with respect to their original party affiliation during sessions preceding confidence vote (Sessions 6 and 7). However, their votes started to change after the vote. At the end their votes are closest to the political opponent of social democrats – ODS.

5 Conclusions

The main aim of this article was to study the development of political views of legislators, who were without political party club affiliation during the selected parliamentary period. In this case, the selected period covered some time before and after confidence vote after 2006 general elections – from October 2006 till May 2007. Two deputies originally from the opposition party – Czech Social Democratic Party (CSSD) - allowed passing the confidence vote. These two deputies had left their political club before this vote; the classification of their votes in session-by-session comparison before and after the confidence vote were a topic of this study. The classification was done using the k-nearest neighbors analysis.

Results of calculations show that both deputies voted with respect to their original party affiliation during sessions preceding confidence vote. However, their votes started to change after the vote; interestingly at the end of observed period their votes were closest to the CSSD long-term opponent political party – Civic Democratic party (ODS).

Acknowledgements

Supported by the grant No. 18-01246S of the Grant Agency GAČR.

References

[1] Freixas, J. & Zwicker, W. S. (2003). Weighted voting, abstention, and multiple levels of approval. *Social Choice and Welfare, 21(3)*, 399-431.
[2] Kramer O. (2013). *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Intelligent Systems Reference Library, vol 51. Springer, Berlin, Heidelberg
[3] Web: https://en.wikipedia.org/wiki/2006_Czech_legislative_election, accessed May 7, 2019.
[4] Web: http://public.psp.cz/en/sqw/hp.sqw?k=193, accessed May 7, 2019.
[5] Web: http://www.psp.cz, accessed May 7, 2019.
[6] Web: http://www.psp.cz/sqw/hlasy.sqw?g=43985&l=cz, accessed May 7, 2019.
[7] Web: http://www.psp.cz/sqw/hlasy.sqw?g=43986&l=cz, accessed May 7, 2019.
[8] Web: http://www.psp.cz/sqw/hlasy.sqw?g=44293&l=cz, accessed May 7, 2019.
[9] Web: http://www.psp.cz/sqw/hlasy.sqw?g=44305&l=cz, accessed May 7, 2019.

Project portfolio designing under risk

Petr Fiala[1], Renata Majovská[2]

Abstract. Project portfolio designing is a dynamic multi-criteria decision-making problem under risk. The paper presents an approach for dynamic project portfolio management based on the Analytic Network Process (ANP) model. The ANP model consists of four basic clusters (projects, resources, criteria, time) with their elements and influences. An important factor of the proposed ANP model is time. The ANP method is suitable for the determination of priorities in network systems where there are different types of dependencies between the elements of the system. In each time period, the portfolio of projects is reviewed in line with the strategic objectives of the organization. Management may decide to initiate new projects, but it is not so easy to predict what new projects will appear in the future. The organization must decide under risk whether to assign all available resources to present proposals or to reserve a portion of the funds unused for some time and wait for better alternatives that may occur later. We propose to use a decision tree with multiple criteria and interactive multi-criteria analysis for solving this problem.

Keywords: project portfolio, multiple criteria, risk, dynamics.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Project management is the discipline of planning, organizing, securing and managing resources to bring about the successful completion of specific project objectives (see [1], [7]). In an accelerating economic world, projects become tools for promoting the objectives of the organization. Project opportunities come in time and it is necessary to decide which will be accepted for creating a dynamic portfolio of projects and which will be rejected. The use of project portfolio management is increasingly becoming a tool for promoting the strategy of the organization, which is a very important role (see [8]). Using of standard methods or trying to design and apply sophisticated methods based on quantitative analysis is possible for portfolio management (see [5]). Selection of project portfolio is a dynamic multi-criteria decision-making problem under risk.

The paper presents an approach for dynamic project portfolio management based on the Analytic Network Process (ANP) model. The ANP model consists of four basic clusters (projects, resources, criteria, time) with their elements and influences. An important factor of the proposed ANP model is time. The ANP method is suitable for the determination of priorities in network systems where there are different types of dependencies between the elements of the system. Time dependent priorities play an increasingly important role in a rapidly changing environment of network systems. Long-term priorities can be based on time dependent comparisons of system elements. Short-term predictions can be based on using of compositional data exponential smoothing.

In each period, the portfolio of projects is reviewed in line with the strategic objectives of the organization. Management may decide to initiate new projects, but also to end of some others that are currently being implemented. Even if the organization has available funds, it is sometimes better to decide not initiate a new project and wait for better one. However, while the set of projects that are currently ready for implementation is clearly defined, it is not so easy to predict what new projects will appear in the future. The organization must decide under risk whether to assign all available resources to present proposals or to reserve a portion of the funds unused for some time and wait for better alternatives that may occur later. We propose to use a decision tree with multiple criteria and interactive multi-criteria analysis for solving this problem. The procedure is in two phases. In the first stage are selected effective strategies. In the second stage, using interactive multi-criteria method selected preferred strategy. The procedure is flexible and can be modified and generalized. The risk attitude of the decision maker

[1] University of Economics, Prague, department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, pfiala@vse.cz.
[2] University of Finance and Administration, Prague, department of Computer Science and Mathematics, Estonska 500, 101 00 Praha 10, Czech Republic, renata.majovska@mail.vsfs.cz.

can be modified. We propose to complete the dynamic ANP model by a decision tree with multiple criteria and interactive multi-criteria analysis for solving the problem of project portfolio management under risk.

2 ANP model

The Analytic Network Process (ANP) is the multi-criteria method (see [10]) that makes it possible to deal systematically with all kinds of dependence and feedback in the performance system. The ANP approach seems to be very appropriate instrument for project portfolio management. Another issue is the appropriate selection of clusters, which would be the basis of the basic model and their fulfilment by system elements. Another specific problem is the creation of sub - networks in the ANP model characterizing the specific important circumstances of the model. The current economic environment is characterized by significant changes. An important problem of the model will be to capture the dynamics that would represent appropriate changes. Time dependent priorities play an increasingly important role in a rapidly changing environment of network systems. Long-term priorities can be based on time dependent comparisons of system elements.

The structure of the ANP model for dynamic project portfolio (DPP) is described by clusters of elements connected by their dependence on one another. A cluster groups elements (projects, resources, criteria, time) that share a set of attributes. At least one element in each of these clusters is connected to some element in another cluster. These connections indicate the flow of influence between the elements (see Fig. 1).

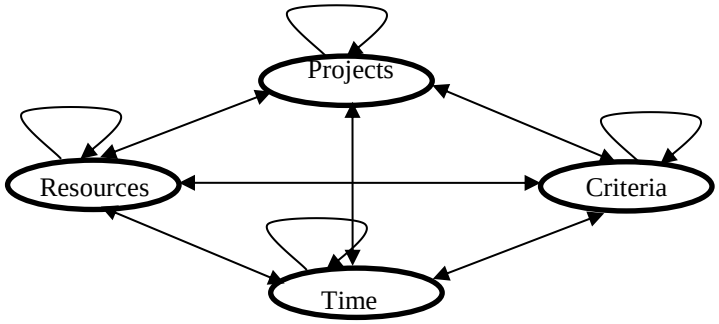


Figure 1 Flows of influence between the elements

The structure of the ANP model for dynamic project portfolio (DPP) is described by clusters of elements connected by their dependence on one another. A cluster groups elements (projects, resources, criteria, time) that share a set of attributes. At least one element in each of these clusters is connected to some element in another cluster. These connections indicate the flow of influence between the elements (see Fig. 2).

The ANP model consists of four basic clusters with their elements and influences:

- Projects: This cluster consists of potential alternatives of projects of which will be selected a dynamic portfolio. There are priorities among projects for inclusion in the portfolio. The cluster has connections to all other clusters.

- Resources: Resources are necessary for the implementation of projects. Main resources are human resources between which are relations important for creating project teams. The cluster has connections to all other clusters.

- Criteria: Projects are evaluated according to criteria which include benefits, opportunities, costs, and risks (BOCR). The cluster has connections to all other clusters.

- Time: Time is measured in discrete units. Elements of other clusters vary in time and theirs values depend on the values in previous time periods.

The basic ANP model is completed by specific sub-networks. The sub-networks are used to model important features of the DPP problem. The most important features in our ANP-based framework for DPP management are captured in sub-networks: dynamic flow of projects, time dependent resources.

- Dynamic flow of projects: Project opportunities come in time and it is necessary to decide which will be accepted for creating a dynamic portfolio of projects and which will be rejected. The sub-network connects clusters: time and projects.

- Time dependent resources: A specific sub-network is devoted to model time dependent amounts of resources. The time dependent amount of resources is given by. The sub-network connects clusters: time, resources and projects.

An important characteristic of project portfolio management is dynamics. Time dependent priorities in the ANP model can be expressed by forecasting using pairwise comparison functions (see [6], [11]). Dynamic extensions of ANP method can work with time-dependent priorities in a networked system. There are two approaches for time-dependent pairwise comparisons: structural, by including scenarios, functional by explicitly involving time in the judgment process.

Judgment matrix in dynamic form is formulated

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \dots & a_{1k}(t) \\ a_{21}(t) & a_{22}(t) \dots & a_{2k}(t) \\ \vdots & \vdots & \vdots \\ a_{k1}(t) & a_{k2}(t) \dots & a_{kk}(t) \end{bmatrix}.$$

For the functional dynamics there are analytic or numerical solutions. The basic idea with the numerical approach is to obtain the time dependent principal eigenvector by simulation (see [11]).

3 Multi-stage decision-making processes

The dynamic approach to project portfolio design is multi-stage. In assessing possible projects, we often encounter a succession of sub-decisions, and the decision-maker's task is to choose from possible sequences that lead to the best target solution. These are multi-stage decision-making processes (see [3]). Decision making takes place in time periods $t = 1, 2, \dots T$. So-called decision trees are used to solve these problems. The two-stage decision-making problems are dealt with in two stages. The first phase is the construction of the decision tree and the second phase is its evaluation.

For the construction of decision trees, a tree graph structure is used, which appropriately models the branching options. When constructing this graph structure of the decision problem, the decision maker creates a decision tree and evaluates its parts in order to find the optimal sequence of decisions. Two kinds of nodes are considered, decision and situation. The decision nodes are marked with a square and the letter *R*, the situation nodes are marked with a circle and the letter *S*. The edges of the tree represent branching of decision and situation possibilities. It starts with a decision node from which origin the edges representing possible decisions a_i. The ends of these edges are situation nodes from which edges represent the possible situations that may occur with conditional probabilities p_j. These decision nodes can be followed by other decision nodes, situation nodes with possible situations, etc. Combination of these basic elements can result in extensive decision trees. End nodes that are no longer linked to other decision and situation nodes are possible ends of sub-order decision sequences that are evaluated.

The decision trees are evaluated in the opposite direction from end edges to the starting decision node. The decision is made by the decision maker who cannot influence the occurrence of the situation and must take into account all situations with their conditional probabilities of occurrence. The selected decision from possible choices is always the one that delivers better ranking. The principle of maximizing expected value is used in the selection. In this way, an optimal sequence of decisions is obtained.

4 Project portfolio management

The most suitable strategy for dynamic project portfolio management is selected using multi-criteria decision trees (see [6], [9]). This approach is chosen because decision trees analyse multi-stage processes and thus capture the dynamic nature of the problem. The multi-criterion variant of decision trees corresponds to the evaluation according to multiple criteria. For their analysis, we will use standard procedures of multi-criteria decision making (see [3]).

A final compromise strategy for the dynamic selection of project portfolios is being sought. This strategy should be effective. An effective multi-criteria strategy is one to which no exists other alternative strategy that would be better at least under one criterion, and not worse under other criteria.

Multi-criteria analysis is two-stage:
• Identify all effective strategies for dynamic portfolio selection.
• An interactive procedure for determining the resulting compromise strategy for dynamic portfolio selection.

The following simple procedure can be applied for the identification of effective strategies:

Step 1: Starting from the last period $t = T$, identify sub-effective strategy for all decision nodes of the period T.

Step 2: Go to the previous period $t = t - 1$.

Step 3: Identify strategies that meet the conditions of effectiveness for each decision node of the period t.

Step 4: If $t > 1$, go to step 2, otherwise the procedure stops.

Number of effective strategies can be large. It is possible to use a simple interactive process between the decision maker and solver for the selection of the final compromise strategy from the set of all effective strategies. In each iteration q, a set of strategies $S(q)$ is analysed and the ideal alternative $\mathbf{H}(q)$ (vector of best values according to each criterion) and the anti-ideal alternative $\mathbf{D}(q)$ (vector of worst values according to each criterion) are determined. The decision maker compares between such values may vary criteria values. The decision maker is asked about the aspiration levels of criteria $\mathbf{A}(q)$, which he would accept as a compromise strategy. If the decision-maker is satisfied with the proposed strategy, the process stops.

Interactive process to determine the final compromise strategy has the following steps:

Step 1: Iteration $q = 1$, the set of all analysed strategies $S(1)$ is equal to the set of all effective strategies.

Step 2: Determine the ideal alternative $\mathbf{H}(q)$ and the anti-ideal alternative $\mathbf{D}(q)$.

Step 3: Decision-maker is asked to accept anti-ideal values. If yes, go to Step 8.

Step 4: The decision-maker is asked to propose the aspiration levels $\mathbf{A}(q)$. If not, go to step 6.

Step 5: The decision-maker enters aspiration levels $\mathbf{A}(q)$ and he determines the corresponding set of acceptable strategies $S(q + 1)$. If $S(q + 1) = \emptyset$, go to step 4, otherwise to step 7.

Step 6: The decision-maker is asked which anti-ideal value is unacceptable for him. A new set of strategies is defined $S(q + 1)$which exceed the unacceptable anti-ideal value.

Step 7: Set $q = q + 1$, go to step 2.

Step 8: The decision-maker is asked which criterion should reach the ideal value. The strategy that maximizes this criterion is the resulting compromise one.

5 Illustrative example

The procedure is illustrated in a simple example with a selection of a project portfolio in two time periods $t = 1, 2$. In the first period, two projects P_1 and P_2 are evaluated. Limited resources allow you to select only one of these projects. In the second period, there are suggestions for other projects P_3 and P_4 with some probability, at the same time both projects with probability 0.2, only project P_3 with probability 0.5 and only project P_4 with probability 0.3. At this stage, the decision maker can select only one of the P_3 and P_4 projects, or no project, due to the limited resources for implementation. The decision tree for this situation is shown in Fig. 2.

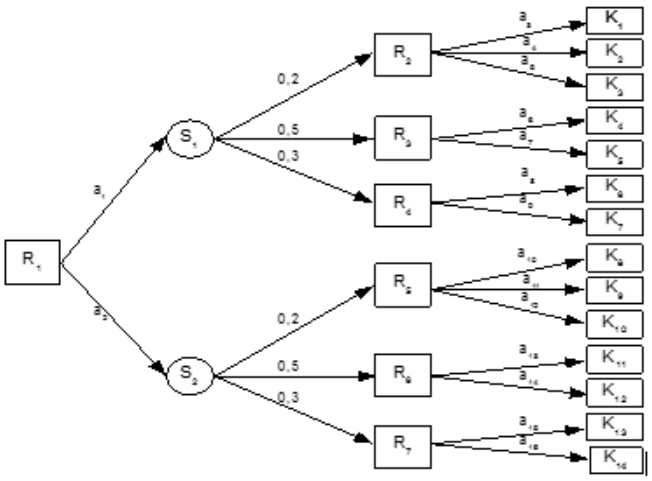


Figure 2 Decision tree for project portfolio selection

Decision node R_1 has two possible decisions, a_1 corresponds to project selection P_1 and a_2 corresponds to project selection P_2. Situation nodes S_1 and S_2 describe three possible states of project proposals, simultaneously both projects P_3 and P_4, only project P_3 and only project P_4 with given probabilities. Decision node R_2 has three possible decisions, a_3 corresponds to project selection P_3, a_4 corresponds to project selection P_4 and a_5 corresponds to the decision not to select any new project. Decision node R_3 has two possible decisions, a_6 corresponds to project selection P_3 and a_7 corresponds to the decision not to select any new project. Decision node R_4 has two possible decisions; a_8 corresponds to the selection of project P_4 and a_9 corresponds to the decision not to select a new project. An analogous tree structure for the situation node S2.

The selected portfolio is rated according to three criteria
f_1: NPV - net present value,
f_2: new market revenue percentage,
f_3: new product revenue percentage.

The estimated criteria values are in Table 1 according to the individual end nodes of the decision tree.

End nodes	f_1	f_2	f_3	End nodes	f_1	f_2	f_3
K_1	70	10	12	K_8	60	15	20
K_2	80	6	15	K_9	70	12	8
K_3	30	5	8	K_{10}	20	7	6
K_4	70	10	12	K_{11}	60	15	20
K_5	30	5	8	K_{12}	20	7	6
K_6	80	6	15	K_{13}	70	12	8
K_7	30	5	8	K_{14}	20	7	6

Table 1 Criteria values in end nodes

The total number of strategies for this decision tree is 24. According to the procedure for identifying effective strategies, we will determine 4 effective strategies according to the expected values (Table 2).

Strategies	f_1	f_2	f_3
s_1: a_1 - a_3 - a_6 - a_8	73	8.8	12.9
s_2: $a_1 - a_4 - a_6 - a_8$	75	8	13.5
s_3: $a_2 - a_{10} - a_{13} - a_{15}$	63	14.1	16.4
s_4: $a_2 - a_{11} - a_{13} - a_{15}$	65	13.5	14

Table 2 Effective strategies

We will use the interactive procedure to determine the final compromise strategy:

1. Iteration q = 1, set of all analyzed strategies $S(1) = \{s_1, s_2, s_3, s_4\}$.
2. $\mathbf{H}(1) = (75; 14.1; 16.4)$, $\mathbf{D}(1) = (63; 8; 12.9)$.
3. The decision maker does not accept anti-ideal values.
4. The decision maker will propose aspiration levels.
5. $\mathbf{A}(1) = (70;8,5;10)$, $S(2) = \{s_1\}$ and the decision maker is satisfied with this strategy.

This strategy selects the P_1 project in the first period and, in the second period, selects the P_3 or P_4 project with the expected criteria values $f_1 = 73$, $f_2 = 8.8$ and $f_3 = 12.9$.

6 Conclusions

For most companies, portfolio selection is a continuous process. Different criteria are usually taken into account when selecting a project portfolio. Deciding when choosing a project portfolio is a risk decision. As a result, this problem can be formulated as a dynamic multi-criteria risk decision.

The proposed portfolio selection procedure respects the characteristics of the problem:

- dynamics,
- multi-criteria,
- risk.

The paper presents an approach for dynamic multi-criteria project portfolio management based on the Analytic Network Process (ANP) model with time dependent priorities. The project portfolio is designed under risk. An interactive method based on multi-criteria decision trees is used for risk evaluation. The procedure has two phases. In the first phase, effective strategies are selected. In the second phase, the preferred strategy is selected using the interactive multi-criteria method. The procedure is flexible and can be modified and generalized. The decision-maker's attitude to risk can be modified, for example, by applying the stochastic dominance rule. Other multi-criteria methods can be used to select the preferred strategy. The selection of the project portfolio is carried out in a number of different application areas, such as supply chain design (see [4]).

Acknowledgements

This paper was written with the support of the grant No. IGA F4/66/2019, Faculty of Informatics and Statistics, University of Economics, Prague. The paper is a result of institutional research project no. 7429/2018/08 supported by University of Finance and Administration, Prague.

References

[1] Fiala, P. (2003). *Projektové řízení - modely, metody, analýzy*. Praha: Professional Publishing.
[2] Fiala, P. (2006). An ANP/DNP analysis of economic elements in today's world network economy. Journal of Systems Science and Systems Engineering, 15(2), 131-140.
[3] Fiala, P. (2008*). Modely a metody rozhodování*. Praha: Oeconomica.
[4] Fiala, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research*, 24(2), 267-281.
[5] Fiala, P. (2018). Project portfolio designing using data envelopment analysis and De Novo optimisation. *Central European Journal of Operations Research*, 26(4), 847-859.
[6] Haimes, Y., Li, D., Tulsiani, V. (1990). Multiobjective decision tree method. *Risk Analysis* 10 (2), 111-129.
[7] Larson, E. W. Gray, C. (2013). *Project Management: The Managerial Process with MS Project*. New York: McGraw-Hill,
[8] Levine, H. A. (2005*). Project Portfolio Management*. San Francisco: Jossey-Bas.
[9] Nowak, M., Nowak, B. (2013). An application of the multiple criteria decision tree in project planning, Procedia Technology 9, 826-835.
[10] Saaty, T. L. (1996). Decision making with Dependence and Feedback: The Analytic Network Process. Pittsburgh: RWS Publications.
[11] Saaty, T. L. (2007). Time dependent decision making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables. Mathematical and Computer Modeling, 46(5), 860-891.

Analysis of TV Advertisments

Lukáš Veverka [1], Lukáš Frýd [2]

Abstract. Evaluation of advertising campaigns is a crucial part of the workload of the marketing department as the company needs to calculate whether the investment in advertising is profitable. Here we propose a new data-driven method for evaluating marketing campaigns, in particular, we study the effect of TV advertisement on the sales of cough medicine on a weekly basis. The exact sales data are hard to obtain as the dataset carries sensitive information, therefore we use Google Trends as a proxy variable. Moreover, as the sales are heavily influenced by the seasonality, the Fourier transformation is used to decompose the time series. The presence of a carry-over effect is also tested since marketing communication might affect sales with a delay. The results of the proposed method are promising and the method allows to determine the return on investment.

Keywords: Data-driven marketing, Fourier transform, Google Trends, Carry-over effect.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Lots of products are sold in seasonal waves – an example of this might a cough medicine which has the biggest peaks during seasons of flu (either at autumn or springtime). Therefore we will compare 2 different attitudes towards the decomposition of time series – classical additive approach and Fourier transformation. We will compare them and specify the advantages and disadvantages of each attitude.

Marketing communication has often a lagged effect which must be evaluated in the model and in case of its presence it must be considered during estimation of parameters as well as the final calculation of return on investment. As it is believed that the lagged effect is geometrically decreasing the usage of Koyck transformation might be a good idea. However, we will show its disadvantages and propose a new method of more flexible estimation of the carry-over effect.

2 Literature review

As mentioned in article *Seasonality in Regression* [**?**] the usage of dummy variables for estimation of seasonality, creates very broad, yet very gradual dips at the seasonal frequencies. In the case of the bigger dataset, the seasonal dummy variables will still remove more power than necessary and might leave unacceptably large local peaks. An alternative to explicitly adding seasonal dummy variables is to approximate seasonality by a function included in the regression. This function which will be periodical and will remove seasonality in both dependent and independent variables just by using time as an input – a combination of cosine and sine functions. The loss in degrees of freedom might be easily computed since the loss is proportional to the number of trigonometric functions. Its ability to detect cycles in the time series is very strong and was used not only in economy but also in biology to confirm cyclicity of Malaria [**?**].

Whether there is a carry-over effect of advertising on sales or not is a very discussed topic among advertisers for ages. The first attempts to prove this effect were based on simple historical studies which were not based on any statistics. When statics was used it was found that the general model which met both goodnesses of fit and accuracy of prediction was Koyck's model of distributed lags [**?**]. This model counts with current and past advertising levels where the coefficients are geometrically decreasing which is perfectly suitable for the carry-over effect. The Koyck's model suppose cumulative infinitely distributed advertising effect and the rate of decrease in range of $(0; 1)$ [**?**].

[1] University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vevl00@vse.cz
[2] University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, lukas.fryd@vse.cz

3 Methodology

3.1 Fourier transformation

Even though Fourier transformation was developed to solve the heat equation, it is used in lots of other disciplines. Primarily it is used for signal processing, electrical engineering as well as in econometrics. In this paper, we will use it for the approximation of seasonality. It works similarly to Taylor polynomial which can approximate a difficult function around a certain point with combining power series. Fourier transformation is based on combining trigonometric functions which can converge to any periodical function. An extreme example is a square or sawtooth wave which are angular. Fourier transformation can converge even to them. The black line in figure **??** represents the above mentioned square wave and it is possible to see that with the growing degree of Fourier series, the resulting function is converging towards it [**?**].

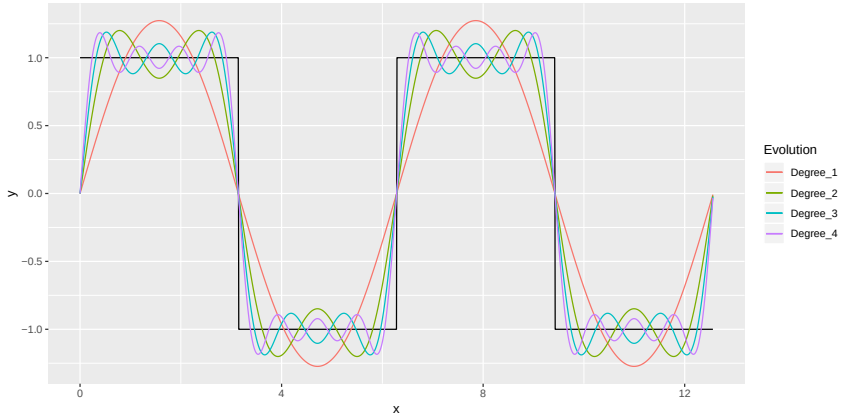


Figure 1: Fourier transformation converging to square wave

Fourier transformation is based on combining trigonometric series. The principal thought is based on assembling anisochronous (i.e. having different frequency) harmonic motion having the same direction with such frequencies so that the resulting function would be periodical. thus $T_1 = nT_n$, where n is integer. The equation for such a function is:

$$f(t) = B_0 + \sum_{n=1}^{\infty} A_n \sin(\omega n t) + \sum_{n=1}^{\infty} B_n \cos(\omega n t), \tag{1}$$

where ω represents how many times the seasonal wave appears in one time period (e.g. $\omega = 2\pi$ in case of one wave in year) and t is linearly increasing vector containing the same number of elements as observations and its values are in range of 0 and length of time periods (e.g. years). Finally, A_n and B_n are coefficients creating spectrum of function [**?**]. These coefficients are then estimated with OLS (ordinary least squares) so the resulting function would approximate the seasonality of given time series.

3.2 Carry-over effect

When analyzing time series of media communication and evaluating its success, it is necessary to take into account a lot of factors which might affect it. It is even possible that the outputs might be unreliable thanks to market development (e.g. sales might be affected not only by investments to media communication or promo activities but also by competitors activity). One of these factors which must be included in the model is a delayed effect of advertisement – *carry-over effect* [**?**].

There might occur two types of this phenomenon. One type is a time gap between the advertisement being on-air and moment of actual sales caused by this ad. The other is a momentum of new customers who were attracted by the ad and will buy products repeatedly in the next periods [**?**].

In this paper, the carry-over effect will be calculated cumulatively. Every time, when the impression (advertisement) occurs, it becomes stronger in two ways. Firstly, there is a tendency to forget the advertisement more slowly then it appears. Then, the effect of previous impressions will combine with the new ones [**?**]. It is possible to describe it by math in this way:

$$E_t = I_t + cE_{t-1}, \tag{2}$$

where E is the intensity of effect caused by advertisement, I is the investment in the current period and c is the power of carry-over effect. Variable c must have its values in the range of $\langle 0; 1\rangle$ (E.g. $c = 0,84$ would mean that 84 % of intensity from the previous period would carry over to the next one). Starting value E_0 will be equal to the first investment I_0. This might lead to a biased result since in case of a massive campaign before the focus period, its effect would not be considered in the model.

In formula (**??**) it is evident that the intensity of the advertising effect decreases geometrically and converges to zero when the investments are over, since $c \in \langle 0; 1\rangle$. The power of the carry-over effect also depends on the frequency of data (or more precisely on the length of one period). In the case of monthly data, the power of the carry-over effect would be smaller than in the case of weekly data. The reason for that is simple – if the delayed effect of advertisement would take for example 3 weeks and we would work with weekly data, the carry-over effect would be strong. However, in the case of monthly data, it would be absorbed by the length of the time period.

4 Results

4.1 Comparison of decomposing the time series

In case we would use one of the easiest decomposition of time series (the combination of moving average and additive seasonal dummy variables), we would get inaccurate results. Actually, they would not be inaccurate but not responding to the reality and their economic interpretation would be difficult. In figure **??** there is clearly visible the problem mentioned in section **??** of unexpected dips (e.g. the one between 14^{th} and 17^{th} week). What is behind the outstanding effect of 15^{th} and 16^{th} week. In the case of longer time series, this fluctuation would be much smaller.

Therefore the smoothed seasonality is much easier to interpret. It is possible to estimate it by using the Fourier transformation mentioned in section **??**. We will create 4 models, where each model will have one more degree of Fourier series. In table **??** it is possible to see the outputs from all models with a different trigonometric combination. For comparing these models it is useful to analyze the variance of the sum of squared residuals (ANOVA). The null hypothesis is that all groups are random samples from the same population or that the models are not significantly different. Concerning table **??**, we can say that a model with 3 degrees of Fourier series is a better estimation of seasonality than the model with 1 degree because of rejection of null hypothesis on the 5% significance level. This result was expected since the adjusted coefficient of determination for the 3 degrees model is higher. As we can see in figure **??**, the green line (standing for the model with 3 degrees) has the highest peak just after the beginning of the year. Probably it is a period with the most incidences of diseases accompanied by cough.

Another great advantage which was not mentioned in the article *Seasonality in Regression* is that Fourier transformation might be used even for years with 53 weeks. That is impossible in case of the additive method (seasonal dummy variables).

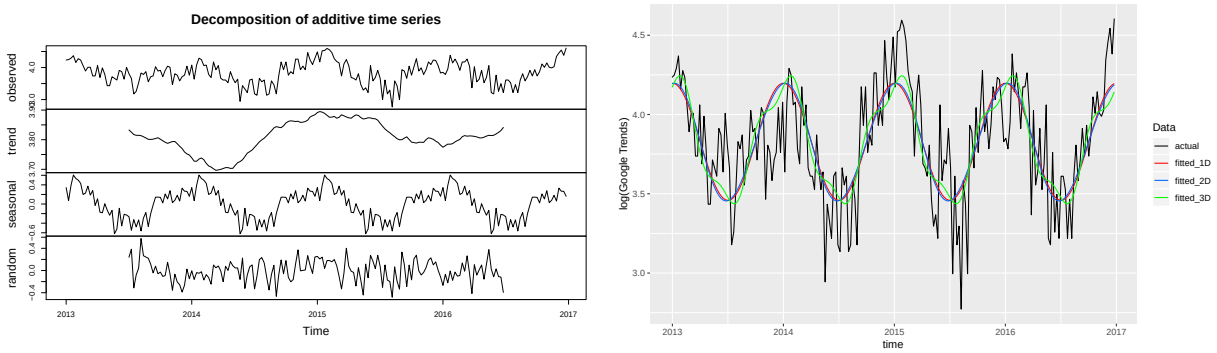
Table 1: ANOVA for models with 1 and 3 degrees of Fourier series

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1 degree	206	12.570				
3 degrees	202	11.893	4	0.677	2.875	0.024**

Notes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4.2 Koyck transformation vs. pre-calculated Carry-over effect

Even though Koyck transformation is popular for estimation of marketing effects, we did not find it suitable for some reasons. Firstly, the main disadvantage is that it is impossible to distinguish different carry-over effects for different communication (e.g. own and competitor's). The reason for that is seen in equation (**??**) where we suppose to have two regressors x_t and z_t. The dynamic marginal effect of x on y would be then $\frac{\partial y_{t+s}}{\partial x_t} = \theta_0 \phi_1^s$ where



(a) Moving average 3 and additive dummy variables

(b) Fourier transformation

Figure 2: Comparison of different types of decomposition

s stands for the infinite lag. The effect of z on y would be $\frac{\partial y_{t+s}}{\partial z_t} = \lambda_0 \phi_1^s$. The Koyck transformation would only cause averaging of both carry-over coefficients as a result of estimation just one parameter (ϕ_1). Another attitude might be VAR models but in that case, we cannot guarantee the geometrically decreasing effect of advertisement.

$$y_t = \delta + \phi_1 y_{t-1} + \theta_0 x_t + \lambda_0 z_t + u_t \tag{3}$$

Another disadvantage of Koyck transformation is that the carry-over effect will be always estimated in the range of (0; 1) and it is impossible to limit it which might be sometimes bad. Let's suppose a campaign of getting a gift conditioned by buying a certain product within a month. It would be irrational to claim that the effect of advertisement is much longer than the month. Therefore the carry-over effect cannot be high either (e.g. 90 % in weekly data would mean almost 60% of the effect in the fifth week – already after the campaign).

Therefore we decided to think of another method – pre-calculated time series of communication. It is based on the thought of creating a new time series (standing for the intensity of advertisement considering even the carry-over effect) from the original media investments. This is based on the iterative computation of equation (**??**) mentioned in section **??**. Then, these vectors will be added to the model instead of the original media investment. Since there are infinity possible models (supposing coefficient for carry-over is a real number), it is impossible to estimate them all and therefore we used an evolutionary algorithm which converged to the best solution. However, there is a possibility of not finding the optimal solution since it estimates the only limited number of models. Also, it might be difficult to determine starting values – the outputs from Koyck transformation might be used though. Our estimated model is:

$$ln(y) = \beta_0 + \beta_{TV} TV + \beta_{C_1} C_1 + \beta_{C_2} C_2 + \sum_{n=1}^{3} \beta_{sin_n} \sin(2\pi t) + \sum_{n=1}^{3} \beta_{cos_n} \cos(2\pi t) + u, \tag{4}$$

where β_0 stands for constant sales of cough medicine, $\beta_{TV}, \beta_{C_1}, \beta_{C_2}$ are the effects of TV communication (either own – TV or competitors' – C_1 and C_1), variables TV, C_1 and C_2 are the pre-calculated time series concerning the carry-over effect. Finally, β_{sin_n} are n parameters for sine in n^{th} degree of Fourier series, the same holds for β_{cos_n} for cosine. The resulting carry-over effects for the best model are shown in table **??** and it is possible to see that the carry-over effects slightly differ from each other (that would not be possible with Koyck transformation).

Table 2: Carry-over effects in the best models

Variable	Carry-over effect
TV advertisements	90%
Competitor's TV	96%
Second competitor's TV	94%

The rating of the model was evaluated only based on the coefficient of determination since the number of variables did not change and the model was manually tested afterward. In case of more complicated models (e.g. including variables selection) it would be necessary to create a model rating function which would be able to evaluate the model based on more aspects than just coefficient of determination (e.g. autocorrelation of the residuals, multicollinearity). The evaluation should take into account also a different number of variables. However, this is not the goal of this paper and it will be part of future research.

As we can see from the results of the final model (see table **??**), there was a significant effect of one of the competitors. Even though the own TV advertisement seems to be insignificant, we have to take into account that the standard errors are affected by weak multicollinearity (tested with variance inflation factor). If the resulting intensity of own TV advertisement (concerning carry-over effect) increases by 1 mil. CZK, the Google trends index would increase by $(e^{0.009} - 1) \times 100 = 0.9$ %.

Table 3: Final model (SE are affected by a weak multicolinearity)

	Dependent variable:
	ln(Google Trends)
TV advertisements (co=90 %)	0.009
	(0.006)
Competitor's TV (co=96 %)	-0.022^{***}
	(0.005)
Second competitor's TV (co=94 %)	0.014^{*}
	(0.010)
Observations	208
R^2	0.613
Adjusted R^2	0.595
F Statistic	34.819^{***} (df = 9; 198)
Notes:	(i) Model included also trigonometric series
	(ii)$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

5 Conclusion

We compared the decomposition of time series based on moving average and seasonal dummy variables and Fourier transformation and we reached the same results as mentioned in the article *Seasonality in Regression*. Seasonal dummy variables cause unacceptably large dips which are hard to rationally explain. Therefore we consider the Fourier transform as a more suitable solution for decomposing seasonal time series. Also, it is not a problem to have a year with 53 weeks when calculating with Fourier transformation. Seasonality was found as the most significant factor which is logical since the cough medicament is mainly a seasonal product. The highest peak is just after the beginning of the year which is probably connected to the most incidences of cough.

The main disadvantages of Koyck transformation which cannot be omitted are the impossibility to limit the coefficient for carry-over effect (e.g. because of the time-limited offer) as well as averaging the effects for all communications and not distinguishing them. However, the results might be very well used as starting values for our suggested method – pre-calculated time series of communication. It is based on an evolutionary algorithm which tests different carry-over effects in a specified model and converges to the best solution. When using this attitude, it must be taken into account that the best solution might not be the optimal solution. By using this method we managed to estimate different carry-over effects for each communication (either own or competitors') and detect the main competitor as well as estimate the effect of own TV advertisement.

Acknowledgements

The work was supported by project F4/19/2019 of Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Sims, C. A. (1974). Seasonality in regression. *Journal of the American Statistical Association*, 69(347), 618–626.

[2] Barth, J. R., & Bennett, J. T. (1974). Cyclical Behavior, Seasonality, and Trend in Economic Time Series. *Nebraska Journal of Economics and Business*, 13(1), 48–69.

[3] Dimitrov, B. D., Valev, D., Werner, R., & Atanassova, P. A. (2013). Cyclic patterns of cerebral malaria admissions in Papua New Guinea for the years 1987-–1996. *Epidemiology & Infection*, 141(11), 2317-2327.

[4] Nerlove, M., Grether, D. M., & Carvalho, J. L. (2014). Analysis of economic time series: a synthesis. Academic Press.

[5] Miloslav, S. (2011). Manažerská ekonomika: 5., aktualizované a doplněné vydání. Grada Publishing as.

[6] Franses, P. H. P. H., & van Oest, R. R. (2004). On the econometrics of the Koyck model (No. EI 2004–07).

[7] Parsons, L. J. (1976). A rachet model of advertising carryover effects. *Journal of Marketing Research*, 13(1), 76–79.

[8] Clarke, D. G. (1976). Econometric measurement of the duration of advertising effect on sales. *Journal of Marketing Research*, 13(4), 345–357.

[9] Tull, D. S. (1965). The carry-over effect of advertising. *Journal of Marketing*, 29(2), 46–53.

Tables

Table 4: Fourier transformation used for approximation of seasonality of dependent variable

	Dependent variable:			
	ln(Google Trends)			
	(1)	(2)	(3)	(4)
$cos(2\pi t)$	0.369***	0.369***	0.369***	0.369***
	(0.024)	(0.024)	(0.024)	(0.024)
$sin(2\pi t)$	0.009	0.009	0.009	0.009
	(0.024)	(0.024)	(0.024)	(0.024)
$cos(4\pi t)$		−0.003	−0.003	−0.003
		(0.024)	(0.024)	(0.024)
$sin(4\pi t)$		0.020	0.020	0.020
		(0.024)	(0.024)	(0.024)
$cos(6\pi t)$			−0.024	−0.024
			(0.024)	(0.024)
$sin(6\pi t)$			0.074***	0.074***
			(0.024)	(0.024)
$cos(8\pi t)$				0.046*
				(0.024)
$sin(8\pi t)$				−0.026
				(0.024)
Constant	3.828***	3.828***	3.828***	3.828***
	(0.017)	(0.017)	(0.017)	(0.017)
Observations	209	209	209	209
R^2	0.532	0.533	0.557	0.568
Adjusted R^2	0.527	0.524	0.544	0.551
Notes:				*p<0.1; **p<0.05; ***p<0.01

Variability of interval-valued data: A refined analysis of an NP-hard problem

Michal Černý [1]

Abstract. This talk contributes to the analysis of computational complexity of statistical procedures for processing interval-valued data. Evaluation of a tight upper bound for the sample variance of an interval-valued dataset is known to be NP-hard. We refine recent results on probabilistic analysis of this problem. Namely, under a natural data-generating stochastic model, we derive new tail bounds for high computation times. The results show that although the problem is NP-hard in the worst case, such cases are encountered very rarely. This is an important message for practice since it shows that the problem is "almost always" efficiently solvable and thus it makes sense to implement the method in statistical packages for interval data.

Keywords: Interval data; nonconvex quadratic programming; average complexity; tail probability bounds

JEL Classification: C61
AMS Classification: 65C60

1 Introduction

Consider a one-dimensional dataset $x = (x_1, \dots, x_n)$. Data analysts are interested in various statistics (such as descriptive statistics, estimators or test statistics) giving information about the data generating distribution. Formally, such a statistic is just a function S of x. The area of partial identification [6] studies what we can infer about the data generating distribution when the dataset x is latent and we have only partial or indirect information about it. Here we consider the following setup, called *interval data model*: We cannot observe x but we can observe upper and lower bounds, i.e. we can observe a compact interval $x_i := [\underline{x}_i, \overline{x}_i]$ such that $\underline{x}_i \leq x_i \leq \overline{x}_i$ a.s. for all $i = 1, \dots, n$. Now $S(x)$ cannot be computed; but we can, at least, get tight bounds for $S(x)$ in the form

$$\overline{S} = \max\{S(x) \mid x_i \in x_i,\ i = 1, \dots, n\}, \tag{1}$$

$$\underline{S} = \min\{S(x) \mid x_i \in x_i,\ i = 1, \dots, n\}. \tag{2}$$

Indeed, now we have $\underline{S} \leq S(x) \leq \overline{S}$ a.s.

The problem of evaluation of (1, 2) for various statistics $S(x)$ has attracted a lot of attention, see e.g. [1, 2] or the reference books [5, 7] and references therein.

1.1 Sample variance

This text is devoted to the sample variance

$$S(x) = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 .$$

Indeed, (1, 2) give a tightest possible range for $S(x)$ when x is unobservable but the intervals $x_1, \dots, x_n$ are.

From the optimization viewpoint, (2) is a convex quadratic programming problem, which can be solved in polynomial time (which is important if n is large). But (1) is NP-hard [4] and this is a serious obstacle.

1.2 Summary of previous results

The basic question is whether NP-hardness of (1) should be interpreted as "the situation is hopeless". In a sense, the answer is no, at least in a stochastic setup. Of course, the worst case always leads to exponential computing time (unless $P = NP$). But the question is whether such hard instances occur frequently or rarely. This is the merit of [3, 4] where the following theorems have been proven.

[1] Department of Econometrics, Faculty of Informatics and Statistics, University of Economics, Winston Churchill Square 4, 13067 Prague, Czech Republic, cernym@vse.cz

Theorem 1 (worst-case complexity of (1)). *Let $x_i^* = \frac{1}{2}(\underline{x}_i + \overline{x}_i)$ be the center of x_i, let $x_i^\Delta = \frac{1}{2}(\overline{x}_i - \underline{x}_i)$ be the diameter of x_i and let $x_i^{1/n} = [x_i^* - x_i^\Delta/n,\ x_i^* + x_i^\Delta/n]$, $i = 1, \ldots, n$. Let $G_n(V,E)$ be the intersection graph of $x_1^{1/n}, \ldots, x_n^{1/n}$, i.e. $V = \{1, \ldots, n\}$ and $\{i,j\} \in E$ iff $x_i^{1/n} \cap x_j^{1/n} \neq \emptyset$ (where $i \neq j$). Problem (1) can be solved in worst-case computation time*

$$O(n \log n + n 2^{\omega_n}), \tag{3}$$

where ω_n is the size of the largest clique in G_n.

If not said otherwise, "log" stands for the natural logarithm.

The method proving the bound (3) is based on a refinement of the algorithm by Ferson et al., see [3, 4] for details. Theorem 1 is good news: it requires computation time exponential in ω_n and not in n. Although in the worst case $\bigcap_{i=1}^n x_i^{1/n} \neq \emptyset$ we have $\omega_n = n$, in many cases it can be $\omega_n \ll n$. Indeed, if we introduce a suitable data-generating stochastic model, we will see that this is the case.

Theorem 2 (average-case complexity of (1)). *(a) Let the centers $x_1^*, \ldots, x_n^*$ form a random sample from a distribution with a Lipschitz continuous cumulative distribution function with Lipschitz constant $L > 0$,*

(b) let the radii $x_1^\Delta, \ldots, x_n^\Delta$ form a random sample from a nonnegative distribution satisfying $\gamma := \mathsf{E}[(x_i^\Delta)^{1+\varepsilon}] < \infty$ for some $\varepsilon > 0$,

(c) let the random samples from (a), (b) be independent.

Then

$$\mathsf{E}\omega_n = O\left(\frac{\log n}{\log\log n}\right) \quad \text{and} \quad \mathsf{E}2^{\omega_n} = O\left(n^{2/\log\log n}\right). \tag{4}$$

Combining (4) with (3) we get average computation time for (1)

$$\begin{aligned} O(\mathsf{E}[n \log n + n \cdot 2^{\omega_n}]) &= O(n \log n + n \cdot \mathsf{E}2^{\omega_n}) \\ &= O(n \log n + n \cdot n^{2/\log\log n}) = O(n \log n + n^{1+(2/\log\log n)} = O(n^{1+\eta}) \end{aligned}$$

with $\eta > 0$ arbitrarily small. This is a surprising result, showing that the NP-hard problem (1) can be on average solved in almost linear time (under the probabilistic model introduced in assumptions (a) – (c)).

Moreover, the analysis [3] shows that cases requiring exponential computation time are extremely rare.

Theorem 3 (tail bound for the distribution of ω_n). *Let $0 < \delta < 1$. For a sufficiently large n it holds* $\Pr[\omega_n \geq \delta n] \leq e^{-n \log\log n}$.

Theorem 3 captures the case when ω_n is linear in n; then the bound (3) is exponential in n and this is the "bad" case requiring exponential computing time. Theorem 3 tell us that such cases occur very rarely, with probability tending to zero faster than exponentially if $n \to \infty$.

The goal of this paper is to refine the proof method of Theorem 3 from [3]. First we show that the bound can be even sharpened to the order $e^{-n \log n}$; this will be done in Corollary 6. Then we introduce probability bounds also on other possible cases: we will estimate the probability that the algorithm works in polynomial time, that it works in a "slightly" superpolynomial time (such as in time $\sim n^{\log n}$) and also the case that it works in "almost" exponential time (such as in time $\sim 2^{\sqrt{n}}$).

2 The main result

First we state a result generalizing the tail bound from Theorem 3. We still assume the probabilistic model (a) – (c) from Theorem 2. The proof of Theorem 4 is postponed to a separate Section 2.2.

Theorem 4. *Let $g(n) \geq \log n$ and $0 \leq c < 1$. If $f(n) \leq c\, g(n) \log g(n)$, then there exists n_0 such that*

$$\Pr[\omega_n \geq g(n)] \leq e^{-f(n)} \quad \text{for } n \geq n_0.$$

2.1 What Theorem 4 tells us about computational complexity of (1)

Before we turn our attention to the proof, let us consider various choices of $f(n)$ and $g(n)$ to see what Theorem 4 tells us about computational complexity of the problem (1). Recall that the crucial term in the complexity bound (3) is $n \cdot 2^{\omega_n}$. To make things easier, we will neglect the term $n \log n$ and identify the computing time with $n \cdot 2^{\omega_n}$.

Corollary 5 (polynomial time complexity)**.** *The probability that the computation time for solving (1) is n^{k+1} or worse is at most $n^{-k(\log k+\log\log n)}$.*

Proof. Choose $g(n) = k\log_2 n$ and $f(n) = (\log 2)\cdot g(n)\log g(n)$. We have

$$\begin{aligned}\Pr[n2^{\omega_n} \geq n^{k+1}] &= \Pr[\omega_n \geq k\log_2 n]\\ &\leq \exp\left[-(\log 2)k\log_2 n\cdot\log k\log_2 n\right]\\ &= \exp\left[-k\log n\cdot\log\left(\frac{k}{\log 2}\log n\right)\right]\\ &\leq \exp\left[-k\log n\cdot\log\left(k\log n\right)\right]\\ &= n^{-k(\log k+\log\log n)}.\end{aligned}$$

□

Corollary 6 (exponential time complexity)**.** *If $0 < \delta \leq 1$, the probability that the computation time for solving (1) is $n2^{\delta n}$ or worse is at most*

$$e^{-\frac{1}{2}\delta n\log\delta n}. \tag{5}$$

Proof. Choose $g(n) = \delta n, f(n) = \frac{1}{2}\delta n\log\delta n$ and proceed similarly as in the proof of Corollary 1. □

Remark. Corollary 2 improves the probability bound from [3]. The bound (5) has the exponent of the order $-n\log n$ (if δ is fixed), while [3] proves the bound with exponent of the order $-n\log\log n$.

Corollary 7 (an intermediate case: a "slightly" superpolynomial time)**.** *The probability that the computation time for solving (1) is $n^{1+\log n}$ or worse is at most $e^{-(\log^2 n)\log\log n}$.*

Proof. Choose $g(n) = \log^2 n$ and $f(n) = \frac{1}{2}\cdot g(n)\log g(n)$. We have

$$\begin{aligned}\Pr[n2^{\omega_n} \geq n^{1+\log n}] &= \Pr[\omega_n \geq \log_2 n^{\log n}]\\ &= \Pr[\omega_n \geq (\log n)\log_2 n]\\ &\leq \Pr[\omega_n \geq \log^2 n]\\ &\leq \exp\left[-\frac{1}{2}\log^2 n\cdot\log\log^2 n\right]\\ &= \exp\left[-\log^2 n\cdot\log\log n\right].\end{aligned}$$

□

Corollary 8 (an intermediate case: almost exponential time)**.** *If $0 < \delta \leq 1$, the probability that the computation time for solving (1) is $n2^{n^\delta}$ or worse is at most $n^{-(\delta/2)\cdot n^\delta}$.*

Proof. Choose $g(n) = n^\delta, f(n) = \frac{1}{2}\delta n\log\delta n$ and proceed similarly as in the proofs of Corollaries 1–3. □

2.2 Proof of Theorem 4

To prove Theorem 4, we will need a tail bound for the binomial distribution.

Lemma 9 ([8])**.** *Let $H(x) = 1 - x + x\log x$. If $B \sim Bi(n,p)$ and $\kappa \geq \mu := np$, then*

$$\Pr[B \geq \kappa] \leq e^{-\mu H(\kappa/\mu)}.$$

Proof of Theorem 4. The first part of the proof mirrors [3] and [8]. We sketch it only shortly for the sake of completeness. First of all, by Lemma 3 of [3], we have

$$p_n := \Pr[\{i,j\} \text{ is an edge of } G_n] = \Pr\left[x_i^{1/n}\cap x_j^{1/n} \neq \emptyset\right] \leq \frac{\alpha}{n} \quad \text{if } i \neq j \tag{6}$$

with

$$\alpha := 1 + \frac{8L(1+\gamma)}{\varepsilon},$$

where G_n has been defined in Theorem 1 and L, γ, ε have been defined in assumptions (a) – (c) of Theorem 2.

Let $\mathbb{I}\{\cdots\}$ be the 0-1 indicator of $\{\cdots\}$. Define

$$U_{ij} := \mathbb{I}\left\{x_i^{1/n} \cap x_i^{1/n} \neq \emptyset\right\}, \quad i \neq j$$

and

$$V_i := \sum_{j \in \{1,\ldots,n\}\ \{i\}} U_{ij}. \tag{7}$$

The important observation is: if $\omega_n \geq k+1$, then for some i, $V_i \geq k$. (Indeed: if, for example, vertices $1, 2, 3, 4$ form a clique of size 4, then $U_{12} = U_{13} = U_{14} = 1$, and thus $V_1 \geq 3$.) So, for a sufficiently large n and $k > \alpha$,

$$\Pr[\omega_n \geq k+1] \leq \Pr[\exists i\ V_i \geq k] \leq n \Pr[V_1 \geq k] \overset{(\star)}{\leq} n \Pr[B \geq k], \tag{8}$$

where $B \sim Bi(n-1, \alpha/n)$. In the above, we have also used the following facts:

- $V_i \sim Bi(n-1, p_n)$ since the summands in (7) are alternatively distributed and independent;
- $p_n \leq \alpha/n$ by (6). Observe that $k > \alpha$ implies that

$$\mathsf{E}V_i = (n-1)\mathsf{E}U_{ij} = (n-1)p_n \leq \mathsf{E}B = (n-1)\frac{\alpha}{n} \leq \alpha < k,$$

and thus $(\star)$ can be seen as an inequality between right tails of two binomial distributions with success probabilities p_n and α/n such that $p_n \leq \alpha/n$.

We will prove that

$$\frac{\Pr[\omega_n \geq g(n)]}{e^{-f(n)}} \longrightarrow 0 \ \text{ with } \ n \to \infty;$$

the claim then easily follows. Let us put $k := g(n)$ and use (8). Then, Lemma 9 with $\mu = \mathsf{E}B = \alpha\frac{n-1}{n}$ implies

$$\begin{aligned}
\frac{\Pr[\omega_n \geq g(n)+1]}{e^{-f(n)}} &\leq \frac{n \Pr[B \geq g(n)]}{e^{-f(n)}} \\
&= \exp\left[f(n) + \log n + \Pr[B \geq g(n)]\right] \\
&\leq \exp\left[f(n) + \log n - \mu H\left(\frac{g(n)}{\mu}\right)\right] \\
&= \exp\left[f(n) + \log n - \mu\left\{1 - \frac{g(n)}{\mu} + \frac{g(n)}{\mu}\log\left(\frac{g(n)}{\mu}\right)\right\}\right] \\
&\leq \exp\left[f(n) + \log n - \mu\left\{-\frac{g(n)}{\mu} + \frac{g(n)}{\mu}\log\left(\frac{g(n)}{\mu}\right)\right\}\right] \\
&= \exp\left[f(n) + \log n + g(n) - g(n)\log\left(\frac{g(n)}{\mu}\right)\right] \\
&= \exp\left[f(n) + \log n + (1+\log\mu)g(n) - g(n)\log g(n)\right] \\
&\leq \exp\left[f(n) + (2+\log\mu)g(n) - g(n)\log g(n)\right] \\
&\leq \exp\left[cg(n)\log g(n) + (2+\log\mu)g(n) - g(n)\log g(n)\right] \\
&= \exp\left[(2+\log\mu)g(n) - (1-c)g(n)\log g(n)\right] \overset{n\to\infty}{\longrightarrow} 0,
\end{aligned}$$

since $1 - c > 0$ and

$$0 < \log\mu = \log\left(\alpha\frac{n-1}{n}\right) = O(1).$$

Indeed, the term $-O(1)\cdot g(n)\log g(n)$ dominantes the term $O(1)\cdot g(n)$ in the limit. □

Acknowledgment. The work was supported by the Czech Science Foundation under project 19-02773S.

References

[1] Antoch, J., Brzezina, M., Miele, R. (2010). A note on variability of interval data. *Computational Statistics* **25**(1), 143–153.

[2] Černý, M., Hladík, M. (2014). The complexity of computation and approximation of the t-ratio over one-dimensional interval data. *Computational Statistics & Data Analysis* **80**, 26–43.

[3] Černý, M., Sokol O., Rada M. (2019) The algorithm by Ferson et al. is surprisingly fast: An NP-hard optimization problem solvable in almost linear time with high probability. *Submitted.* Preprint available from: `https://arxiv.org/abs/1905.07821`.

[4] Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., Aviles, M. (2005). Exact bounds on finite populations of interval data. *Reliable Computing* **11**(3), 207–233.

[5] Kreinovich, V., Lakeyev, A. V., Rohn, J., Kahl, P. (2013). *Computational Complexity and Feasibility of Data Processing and Interval Computations.* Springer.

[6] Manski, C. F. (2003). *Partial Identification of Probability Distributions.* Springer.

[7] Nguyen, H. T., Kreinovich, V., Wu, B., Xiang, G. (2012). *Computing Statistics under Interval and Fuzzy Uncertainty.* Springer.

[8] Penrose, M. (2003). *Random Geometric Graphs.* Oxford University Press.

Desirable Properties of Weighting Vector in Pairwise Comparisons Matrix

Jaroslav Ramík[1]

Abstract. The aim of this paper is to show a new categorization of inconsistent pairwise comparisons matrices (PCM) proposed with respect to a satisfaction/violation of selected PCM properties, such as the fundamental selection (FS) condition, preservation of order preference (POP) condition, preservation of order of intensity of preference (POIP) condition, and reliability priority condition (RP). A new non-linear optimization problem for finding the weights (i.e. priority vector) satisfying the aforementioned conditions is proposed such that the distance between the given PCM matrix and the ratio matrix composed of the weights function is minimized. Moreover, we presented important examples of alo-groups where the above optimization problem for finding the desirables weights can be solved by standard optimization methods.

Keywords: AHP, pairwise comparisons, pairwise comparisons matrix, ranking of variants, vector of weights, unimodality.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Pairwise comparisons matrix (PCM) plays an essential role in multi-criteria decision making (MCDM), particularly in the problem of ranking n variants, [8]. A popular method for finding the priority vector (i.e. vector of weights) is to find the distance of the given PCM to the closest consistent matrix which is a ratio matrix composed of the weights. In this paper, we show, that such distance function is not necessarily semistrictly quasiconvex, (which is a necessary and sufficient condition for unimodality), [12], moreover, under some assumptions it is proven that each local minimum of the function is the global minimum, i.e., in that sense the objective function is unimodal. Unimodalitity enables finding optimal solutions by standard convex optimization methods, [4]. In order to emphasize the general nature of the relationship between priority vectors and inconsistency of PCMs we decided to use pairwise comparisons based on linearly ordered abelian groups, [5], [6].

2 Preliminaries

Here, we deal with the popular problem of ranking the set of alternatives $\mathscr{C} = \{c_1, c_2, ..., c_n\}$ based on some information of the PC matrix $\mathbf{A} = [a_{ij}]$. The elements a_{ij} of *PC* matrix $\mathbf{A} = [a_{ij}]$ belong to $\mathscr{G} = (G, \odot, \leq)$, an alo-group which will be defined bellow.

An *abelian group* is a set, G, together with an operation $\odot$ (read "*odot*") that combines any two elements $a, b \in G$ to form another element in G denoted by $a \odot b$, see [3, 5]. The symbol $\odot$ is a general placeholder for some concretely given operation. $(G, \odot)$ satisfies the following requirements known as the *abelian group axioms*, particularly: *commutativity*, *associativity*, there exists an *identity element* $e \in G$ and for each element $a \in G$ there exists an element $a^{(-1)} \in G$ called the *inverse element to a*. The *inverse operation* $\div$ to $\odot$ is defined for all $a, b \in G$ as follows: $a \div b = a \odot b^{(-1)}$. Note that the inverse operation is not necessarily associative.

An ordered triple $(G, \odot, \leq)$ is said to be an *abelian linearly ordered group*, *alo-group* for short, if $(G, \odot)$ is a group, $\leq$ is a linear order on G, and for all $a, b, c \in G$: $a \leq b$ implies $a \odot c \leq b \odot c$. In other words, $\odot$ respects $\leq$.

If $\mathscr{G} = (G, \odot, \leq)$ is an alo-group, then G is naturally equipped with the order topology induced by $\leq$ and $G \times G$ is equipped with the related product topology. We say that $\mathscr{G}$ is a *continuous alo-group* if $\odot$ is continuous on $G \times G$. The *inverse operation* $\div$ to $\odot$ is defined for all $a, b \in G$ as follows: $a \div b = a \odot b^{(-1)}$. By definition, an alo-group $\mathscr{G}$ is a lattice ordered group. Hence, there exists $\max\{a, b\}$, for each pair $(a, b) \in G \times G$. Nevertheless, a nontrivial alo-group $\mathscr{G} = (G, \odot, \leq)$ has neither the greatest element nor the least element. Because of the associative property, the operation $\odot$ can be extended by induction to n-ary operation.

[1] Silesian University Opava, School of Business Administration in Karvina, University Sq. 1934/3, 733 40 Karvina, Czechia, ramik@opf.slu.cz

$\mathscr{G} = (G, \odot, \leq)$ is *divisible* if for each positive integer n and each $a \in G$ there exists the (n)-th root of a denoted by $a^{(1/n)}$, i.e., $\left(a^{(1/n)}\right)^{(n)} = a$. The function $\|.\| : G \to G$ defined for each $a \in G$ by

$$\|a\| = \max\{a, a^{(-1)}\} \tag{1}$$

is called a $\mathscr{G}$*-norm*. We shall assume that G is a divisible and continuous alo-group. Then G is an open interval in $\mathbb{R}$. The operation $d : G \times G \to G$ defined by $d(a,b) = \|a \div b\|$ for all $a, b \in G$ is called a $\mathscr{G}$*-distance*. Next, we present the well known examples of alo-groups; for more details see also [5], or, [10].

Examples of alo-groups.
1. Additive alo-group $\mathscr{R} = (\mathbb{R}, +, \leq)$ is a continuous alo-group with:
$e = 0,\ a^{(-1)} = -a,\ a^{(n)} = n \cdot a.$

2. Multiplicative alo-group $\mathscr{R}_+ = (\mathbb{R}_+, \cdot, \leq)$ is a continuous alo-group with: $e = 1,\ a^{(-1)} = a^{-1} = 1/a,\ a^{(n)} = a^n$. Here, by '$\cdot$' we denote the usual operation of multiplication.

3. Fuzzy additive alo-group $\mathscr{R}_a$=$(\mathbb{R}, +_f, \leq)$, see [10], is a continuous alo-group with:

$$a +_f b = a + b - 0.5,\ e = 0.5,\ a^{(-1)} = 1 - a, a^{(n)} = n \cdot a - \frac{n-1}{2}.$$

4. Fuzzy multiplicative alo-group $]\mathbf{0}, \mathbf{1}[_{\mathbf{m}}$=$(]0,1[, \bullet_f, \leq)$, see [5], is a continuous alo-group with:

$$a \bullet_f b = \frac{ab}{ab + (1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a.$$

Remark 1. Usually, the *PC* method is used with a multiplicative *PC* matrix, i.e., with multiplicative alo-group, see e.g. [13, 9]. Then the relative importance of an alternative is multiplied with the relative importance of the other alternatives when considering a chain of alternatives.

The $n \times n$ (square) matrix $\mathbf{A} = [a_{ij}]$, where $a_{ij} \in G$, is called the *pairwise comparison matrix (PCM)*. The matrix $\mathbf{A} = [a_{ij}]$ is *reciprocal* if

$$a_{ij} = a_{ji}^{(-1)} \ \ \forall i, j \in \{1, \ldots, n\}. \tag{2}$$

Hereinafter, it is assumed that *all pairwise comparison matrices are reciprocal.* The matrix $\mathbf{A}$ is *consistent* if

$$a_{ik} = a_{ij} \odot a_{jk} \ \ \forall i, j, k \in \{1, \ldots, n\}. \tag{3}$$

The result of the pairwise comparisons method based on a PC matrix $\mathbf{A} = [a_{ij}]$ is a rating of set $\mathscr{C}$ of alternatives - a mapping that assigns real values to the alternatives. It is a function $w : \mathscr{C} \to G$ that assigns to every alternative from $\mathscr{C} = \{c_1, c_2, ..., c_n\}$ a real number from alo-group G. Here, $w(c)$ represents the ranking value for $c \in \mathscr{C}$. The w function is usually written in the form of a vector of *weights*, i.e., $\mathbf{w} = (w(c_1), ..., w(c_n)) \in G^n$, and is called the *priority vector*. Also, we say that priority vector $\mathbf{w}$ is associated with the PC matrix $\mathbf{A}$, or, alternatively, that the priority vector $\mathbf{w}$ is generated by a priority generating method based on the PC matrix $\mathbf{A}$. Usually, we require that the priority vector is *normalized*, i.e. $\bigodot_{i=1}^{n} w(c_i) = e$. The following proposition gives a characterization of each consistent PC matrix by its priority vector, see e.g. [13].

Proposition 1. *Let* $\mathbf{A} = [a_{ij}]$ *be a consistent pairwise comparison matrix. Then there exists a unique priority vector* $\mathbf{w} = (w_1, ..., w_n) \in G^n$ *associated to* $\mathbf{A}$ *satisfying*

$$a_{ij} = w_i \div w_j, \forall i, j. \tag{4}$$

3 Desirable properties of the priority vector

Some inconsistent pairwise comparisons matrices violate the so called fundamental selection (FS) condition of multiple criteria decision making: the 'best' alternative is selected from the set of non-dominated alternatives, while this set is non-empty. Inconsistent PCMs that violate FS condition should be viewed as logically flawed and should not be used for the derivation of weights of alternatives or other objects. Other PCMs may violate the preservation of order of preferences (the so called POP condition), or preservation of the intensity of preference (the so called POIP condition), see Bana e Costa and Vansnick [1].

Furthermore, a new non-linear optimization problem/s is/are proposed for generating a priority vector (weights of alternatives, criteria, or other objects). The method is designed to find a priority vector so that all four aforementioned logical properties are satisfied, hence providing more logical solution than the eigenvalue (EV), or the geometric mean (GM) methods.

Definition 1. Let $\mathbf{A} = [a_{ij}]$ be the PC matrix based on the set of alternatives $\mathscr{C} = \{c_1, c_2, ..., c_n\}$ with elements a_{ij} from alo-group $\mathscr{G}$. We say that an alternative c_i *dominates alternative* c_j, or, equivalently, that an alternative c_j *is dominated by alternative* c_i, if

$$a_{ij} > e, \tag{5}$$

where e is the identity element of $\mathscr{G}$. If a given alternative is not dominated by any other alternative, then such alternative is called the *non-dominated alternative*. The set of all non-dominated alternatives in $\mathscr{C}$ with respect to matrix $\mathbf{A}$ is denoted by $ND(\mathbf{A})$. By (5) we obtain

$$ND(\mathbf{A}) = \{c_j \in \mathscr{C} | \text{ there is no } i \in \{1, ..., n\} : a_{ij} > e\}. \tag{6}$$

We say that the *fundamental selection (FS) condition* is satisfied with respect to $\mathbf{A}$ and $\mathbf{w} \in G^n$, if the maximal weight of the priority vector is associated to a non-dominated alternative.
Equivalently, we say that $\mathbf{w}$ satisfies the FS condition with respect to $\mathbf{A}$, if for some $i^* \in \{1, ..., n\}$:

$$c_{i^*} \in ND(\mathbf{A}) \text{ and } w(c_{i^*}) = \max\{w(c_j) | j \in \{1, ..., n\}\}. \tag{7}$$

Alternatively, we say that $\mathbf{A}$ satisfies FS condition with respect to $\mathbf{w}$.
A PC matrix $\mathbf{A}$ is said to satisfy *preservation of order preference condition (POP condition) with respect to priority vector* $\mathbf{w}$ if

$$a_{ij} > e \Rightarrow w_i > w_j. \tag{8}$$

A PC matrix $\mathbf{A}$ is said to satisfy *preservation of order intensity preference condition (POIP condition) with respect to vector* $\mathbf{w}$ if

$$a_{ij} > e, a_{kl} > e, \text{ and } a_{ij} > a_{kl} \Rightarrow w_i \div w_j > w_k \div w_l. \tag{9}$$

A PC matrix $\mathbf{A}$ is said to satisfy *reliable preference (RP) condition with respect to priority vector* $\mathbf{w}$ if

$$a_{ij} > e \Rightarrow w_i > w_j, \tag{10}$$

$$a_{ij} = e \Rightarrow w_i = w_j. \tag{11}$$

From (10) in the above definition it is clear that if a PC matrix $\mathbf{A}$ satisfies RP condition with respect to priority vector $\mathbf{w}$, then $\mathbf{A}$ satisfies POP condition with respect to priority vector $\mathbf{w}$.

Remark 2. It is well known from theory of graphs, see e.g. [2], that if there is no cycle of pairs of indexes: $(k_1, k_2), (k_2, k_3), ..., (k_{n-1}, k_n), (k_n, k_1)$, where $k_i \in \{1, ..., n\}$, such that $\{k_1, k_2, ..., k_n\}$ is a permutation of $\{1, 2, ..., n\}$ with $a_{k_i, k_{i+1}} > e, i \in \{1, ..., n-1\}$, and $a_{k_n, k_1} > e$, then $ND(\mathbf{A})$ is non-empty.

Remark 3. Let $\mathbf{A} = [a_{ij}]$ be a consistent pairwise comparison matrix, and let $\mathbf{w} = (w_1, ..., w_n)$ be a priority vector associated to $\mathbf{A}$ satisfying (4). Then it is obvious that FS, POP, POIP and RP conditions are satisfied. From Definition 1 it is clear that if POP condition is satisfied, then FS condition is also satisfied. Moreover, for a consistent pairwise comparison matrix, it is well known, see e.g. [13], that the priority vector satisfying (4) can be generated either by EVM, or, by GMM.

4 Derivation of the priority vector

4.1 Problem (P 0)

It was shown in [11], that the derivation of a priority vector by EV or GM methods from an inconsistent pairwise comparison matrix may result in violating desirable conditions FS, POP, or POIP. Therefore, an alternative approach to the derivation of a priority vector for PCMs may be formulated in terms of satisfaction of FS, POP and POIP conditions.

Let $\mathbf{A} = [a_{ij}]$ be a PC matrix. Based on this PCM, we define the following three sets of indexes $I^{(1)}, I^{(2)}, I^{(4)}$:

$$I^{(1)}(\mathbf{A}) = \{(i, j) | i, j \in \{1, \ldots, n\}, a_{ij} = e\}, \tag{12}$$

$$I^{(2)}(\mathbf{A}) = \{(i,j)|i,j \in \{1,\ldots,n\}, a_{ij} > e\}, \tag{13}$$

$$I^{(4)}(\mathbf{A}) = \{(i,j,k,l)|i,j,k,l \in \{1,\ldots,n\}, a_{ij} > e, a_{kl} > e, a_{ij} > a_{kl}\}. \tag{14}$$

Let $\mathbf{w} = (w_1,...,w_n) \in G^n$ be a priority vector associated to $\mathbf{A}$. An *error function*, $\mathbf{F_A}$, of $\mathbf{A}$ and $\mathbf{w}$ is defined as follows

$$\mathbf{F_A} : \mathbf{w} = (w_1,\ldots,w_n) \in G^n \longrightarrow \mathbf{F_A}(\mathbf{w}) \in G,$$

$$\mathbf{F_A}(\mathbf{w}) = \max\{\|a_{ij} \div (w_i \div w_j)\| | i,j \in \{1,...,n\}\}. \tag{15}$$

The problem (P 0) of finding a priority vector satisfying conditions FS, POP, POIP, and RP, can be formulated in terms of the following optimization problem, where $\mathbf{A} = [a_{ij}]$ is a given PC matrix and $\mathbf{w} = (w_1,\ldots,w_n)$ is an unknown priority vector with variables $w_1,\ldots,w_n$:
(P 0)

$$\mathbf{F_A}(\mathbf{w}) \longrightarrow \min; \tag{16}$$

subject to

$$\bigodot_{k=1}^{n} w_k = e, w_r \in G \ \ \forall r, \tag{17}$$

$$w_r = w_s \ \ \forall (r,s) \in I^{(1)}(\mathbf{A}), \tag{18}$$

$$w_r > w_s \ \ \forall (r,s) \in I^{(2)}(\mathbf{A}), \tag{19}$$

$$w_r \div w_s > w_t \div w_u \ \ \forall (r,s,t,u) \in I^{(4)}(\mathbf{A}). \tag{20}$$

The objective function in (16) minimizes the distance between the elements of PC matrix $\mathbf{A}$ and corresponding elements of "ratio" PCM $\mathbf{W} = [w_i \div w_j]$, measured by distance function δ. By (19), the POP condition is secured and by (20) the POIP condition is satisfied, whereas by (18), (19) the RP condition is satisfied.

4.2 Transformation of problem (P 0) to (P ε)

Unfortunately, (P 0) is not a standard optimization problem, appropriate for solving by existing numerical methods. Here, variables w_i are asked to be strictly positive and some inequality constraints, (19), (20), are strict, hence the set of feasible solution is not closed. That is why we transform the problem into a more convenient one. Let $\varepsilon > e$.
(P ε)

$$\mathbf{F_A}(\mathbf{w}) \longrightarrow \min; \tag{21}$$

subject to

$$\bigodot_{k=1}^{n} w_k = e, w_r \in G \ \ \forall r, \tag{22}$$

$$w_r = w_s \ \ \forall (r,s) \in I^{(1)}(\mathbf{A}), \tag{23}$$

$$w_r \geq w_s \odot \varepsilon \ \ \forall (r,s) \in I^{(2)}(\mathbf{A}), \tag{24}$$

$$w_r \div w_s \geq w_t \div w_u \odot \varepsilon \ \ \forall (r,s,t,u) \in I^{(4)}(\mathbf{A}). \tag{25}$$

Here, strict inequalities in (P 0) have been changed to the non-strict ones by introducing a sufficiently "small" constant parameter $\varepsilon > e$. The following proposition says that both problems, i.e. (P 0) and (P ε), are in some sense equivalent. The proof is evident.

Proposition 2. *Problem (P 0) has a feasible solution* $\mathbf{w}$ *if and only if there exists* $\varepsilon > e$ *such that* $\mathbf{w}$ *is a feasible solution of (P* ε*).*
Moreover, if (P 0) has an optimal solution $\mathbf{w}^*$ *then there exists* $\varepsilon > e$ *such that* $\mathbf{w}^*$ *is an optimal solution of (P* ε*).*

In problem (P ε) we distinguish the following 4 optimization problems depending on the formulation of the objective function (21) as well as constraints (22) - (25), i.e. "nested sets" of feasible solutions. We shall consider the following variants of optimization problems:

(I) Minimize the objective function (21), subject to (22). The optimal solution is denoted by $\mathbf{w}^{(\mathbf{I})}$. FS, POP, POIP and RP conditions are not necessarily satisfied.
(II) Minimize the objective function (21) subject to constraints (22), (24). The optimal solution is denoted by $\mathbf{w}^{(\mathbf{II})}$. POP condition is satisfied; then by Definition 1, and (7), FS condition is also satisfied. However, POIP and also RP condition is not necessarily satisfied.

(III) Minimize the objective function (21) subject to constraints (22), (24), and (25). The optimal solution is denoted by $\mathbf{w}^{(\mathbf{III})}$. Here, FS, POP and POIP conditions are satisfied, however, RP condition is not necessarily satisfied.
(IV) Minimize the objective function (21) subject to constraints (22), (23), (24), and (25). The optimal solution is denoted by $\mathbf{w}^{(\mathbf{IV})}$. Here, FS, POP, POIP and RP conditions are satisfied.

4.3 Solving problem (P ε)

Notice that the set of feasible solutions of (P ε), (21) - (25), could be empty, e.g. for problems (II), and/or (III), see bellow. Even for nonempty set of feasible solutions the optimal solution of the corresponding optimization problems of (I) - (IV) do not exist, as the set of feasible solutions is not secured to be closed and/or bounded and the objective function need not be convex.

On the other hand, if the optimal solution $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$ of some problems of (I) - (IV) of (P ε) exists, then FS, POP, POIP and RP conditions are satisfied depending on the properties of feasible solution sets. Hence, $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$ is an appropriate "optimal" priority vector associated to $\mathbf{A}$ satisfying the required properties.

Proposition 3. *Let $\mathbf{A} = [a_{ij}]$ be a consistent pairwise comparison matrix. Then there is a unique optimal solution $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$ of (P 0) satisfying:*

$$a_{ij} = w_i^* \div w_j^* \ \ \forall i, j, \tag{26}$$

such that FS, POP, POIP and RP conditions are met.

Choosing different alo-groups, e.g. additive alo-group $\mathscr{R} = (\mathbb{R}, +, \leq)$, or, multiplicative alo-group $\mathscr{R}_+ = (\mathbb{R}_+, \cdot, \leq)$, we obtain individual optimization problems of (P ε). For some alo-groups, e.g. multiplicative alo-group $\mathscr{R}_+$, variant (I), (II), or (IV) of problem (P ε), we obtain optimization problems solvable by standard optimization methods, e.g. gradient or penalty methods, such that each local minimum is always global, see [4].

4.4 Examples of problems (P ε)

Finally, we present two examples of optimization problems (P ε) with additive alo-group $\mathscr{R}$, and, multiplicative alo-group $\mathscr{R}_+$.

1. Let $\mathscr{R} = (\mathbb{R}, +, \leq)$ be the additive alo-group. Then the problem (21) - (25) has the following form:
(PA ε)
$$\mathbf{F}_{\mathbf{A}}(\mathbf{w}) = \max\{|a_{ij} - (w_i - w_j)| | \forall i, j\} \longrightarrow \min; \tag{27}$$
subject to
$$\sum_{k=1}^{n} w_k = 0, w_r \in \mathbb{R} \ \ \forall r, \tag{28}$$
$$w_r = w_s \ \ \forall (r,s) \in I^{(1)}(\mathbf{A}), \tag{29}$$
$$w_r - w_s \geq \varepsilon \ \ \forall (r,s) \in I^{(2)}(\mathbf{A}), \tag{30}$$
$$w_r - w_s - (w_t - w_u) \geq \varepsilon \ \ \forall (r,s,t,u) \in I^{(4)}(\mathbf{A}). \tag{31}$$
Problem (PA ε) can be solved as a linear programming problem by simplex method, or, by an interior point method.
2. Let $\mathscr{R}_+ = (\mathbb{R}_+, \cdot, \leq)$ be the multiplicative alo-group. Then the problem (21) - (25) has the following form:
(PM ε)
$$\mathbf{F}_{\mathbf{A}}(\mathbf{w}) = \max\{\max\{\frac{a_{ij} w_j}{w_i}, \frac{w_i}{a_{ij} w_j}\} | \forall i, j\} \longrightarrow \min; \tag{32}$$
subject to
$$\prod_{k=1}^{n} w_k = 1, w_r \geq \varepsilon \ \ \forall r, \tag{33}$$
$$w_r = w_s \ \ \forall (r,s) \in I^{(1)}(\mathbf{A}), \tag{34}$$
$$\frac{w_r}{w_s} \geq \varepsilon \ \ \forall (r,s) \in I^{(2)}(\mathbf{A}), \tag{35}$$
$$\frac{w_r}{w_s} \geq \frac{w_t}{w_u}.\varepsilon \ \ \forall (r,s,t,u) \in I^{(4)}(\mathbf{A}). \tag{36}$$

Notice, that strict inequalities in (P 0) have been changed to the non-strict ones by introducing a sufficiently "small" constant $\varepsilon > e$. Moreover, by an appropriate substitution of variables w_k, product of variables (33) may be transformed to the sum of variables. Here, problem (32) - (35) can be solved as a convex programming problem by an interior point method. However, constraint (36) is not convex, therefore, it requires a special treatment, e.g. an approximation by convex constraints. Then each local minimum is always global one.

5 Conclusion

In this paper, we proposed a new categorization of inconsistent pairwise comparisons matrices with respect to a satisfaction/violation of selected PCM properties, such as the fundamental selection (FS) condition, preservation of order preference (POP) condition, preservation of order of intensity of preference (POIP) condition, and, reliability priority condition (RP). A new non-linear optimization problem for finding the weights (i.e. priority vectors) satisfying the aforementioned conditions is proposed such that the distance function between the given PCM matrix and the ratio matrix composed of the weights is minimized. Moreover, we presented important examples of alo-groups where the above optimization problem for finding the desirables weights can be solved by standard optimization methods, e.g. interior point methods.

Acknowledgements

This research has been supported by GACR project No. 18-01246S.

References

[1] Bana e Costa, C. A., Vansnick, J. (2008). A critical analysis of the eigenvalue method udes to derive priorities in AHP. *European Journal of Operational Research*,187, 3, 1422–1428.

[2] Bollobas, J. (2002). *Modern graph theory.* Springer-Verlag: Berlin-Heidelberg, New York.

[3] Bourbaki, N. (1998). *Algebra II.* Springer Verlag, Heidelberg-New York-Berlin.

[4] Boyd, S., Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paolo, Delhi.

[5] Cavallo, B., D'Apuzzo, L. (2009). A general unified framework for pairwise comparison matrices in multicriteria methods. *International Journal of Intelligent Systems* 24, 4, 377–398.

[6] Cavallo, B., D'Apuzzo, L. (2012). Deriving weights from a pairwise comparison matrix over an alo-group. *Soft Computing* 16, 353–366.

[7] Cavallo, B., D'Apuzzo, L. (2016). Ensuring reliability of the weighting vector: Weak consistent pairwise comparison matrices. *Fuzzy Sets and Systems* 296, 21–34.

[8] Kou, G., Ergu, D., Lin, C. S.; Chen, Y. (2016). Pairwise comparison matrix in multiple criteria decision making. *Technological and Economic Development of Economy* 22, 5, 738–765.

[9] Kulakowski, K., Mazurek, J., Ramik, J. and Soltys, M. (2019). When is the condition of order preservation met? *European Journal of Operat. Res.* 277, 248–254.

[10] Ramik, J. (2015). Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences* 297, 236–253.

[11] Ramik, J. (2017). Ranking Alternatives by Pairwise Comparisons Matrix and Priority Vector. *Scientific Annals of Economics and Business,* 64 (SI), 85–95.

[12] Rubinov, A. (2000). *Abstract Convexity and Global Optimization.* Springer Science + Business Media B.V., Boston-Dordrecht-London.

[13] Saaty, T. L. (1980). *Analytic Hierarchy Process*. McGraw-Hill, New York.

Experimental finance: the Gender Differences in the Disposition Effect Bias

Hana Dvořáčková[1]

Abstract. The disposition effect has been described in the stock-investing context as a behavioral tendency of investors to hold on to losing stocks for too long and sell winning stocks too soon. In this paper it is examined whether the disposition effect bias is different for males and females. The experimental data set was collected by Jochec during years 2009 to 2015, students were trading under standardized rules. In this paper the holding periods of profitable and unprofitable trades were tested and compared. Significant difference in the length of trades and also in the risk approach of males and females.

Keywords: behavioral finance, disposition effect, FX trading, financial modelling, experimental finance

JEL Classification: C12, G02
AMS Classification: 91F99

1 Introduction

The disposition effect has been described in the stock-investing context as a behavioral tendency of investors to hold on losing stocks for too long and sell winning stocks too soon. It is one implication of the Kahneman and Tversky´s (1979) prospect theory to investment. According to them, a person who has not made peace with his losses tend to accept risk and gambles, which would not be otherwise acceptable. According to Odean (1998) the most obvious explanations—explanations based on informed trading, rebalancing, or transaction costs—fail to capture important features of the data. For the purpose of that paper, records of 10 000 accounts at a large discount brokerage house were analyzed to prove the tendency of investors to hold losing investments too long and sell winning investments too soon. Afterwards Shefrin and Statman (1985) published first paper labeling this behavioral effect as the disposition effect.

Over time many studies focused on the trader's behavior and the disposition effect were published. For instance Barberis and Xiong (2009) investigated whether the prospect theory preferences can predict a disposition effect; Kaustia (2010) included chapter a focused on the disposition effect into the book Behavioral Finance: Investors, Corporations, and Markets. In the Choe and Eom (2009) it was examined whether the disposition effect exists in the Korean stock index futures market. Those authors also found strong evidence for the disposition effect and explained it in terms of investor characteristics. Chen, Kim, Nofsinger and Rui (2007) studied the investment decision making in an emerging market. According to them Chinese investors, besides other, tend to sell stocks that have appreciated in price, but not those that have depreciated in price, which is consistent with the disposition effect acknowledging gains but not losses. Results of Marciukaityte and Szewczyk (2012) are in accordance with the proposition that the disposition effect increases the supply of winning stocks and depresses their prices. Contrary, Locke and Mann (2005) found no evidence of any contemporaneous measurable costs associated with the disposition effect.

The aim of this paper is to examine the difference in tendency to succumb to the disposition effect with regard to the gender.

2 Methodology

To show the significance of findings the standard independent two sample t-test for the difference between two means, which has been chosen based on the f-test results. The Levene´s test is used for the testing of the equality of variances and based on the results the choice for the t-test is made. Moreover, the one-way ANOVA test is used for the conclusion of the tested hypotheses. Calculations were done in the program SPSS. Following subchapters are focused on the data collection methodology.**OANDA Trading Game and Data Analysis**

[1] VŠB-TUO, Faculty of Economic, Department of Finance, Sokolská 33, Ostrava, 702 00, hana.dvorackova@vsb.cz

The dataset used in this research was collected by Jochec (Dvořáčková and Jochec 2018) from 2009 to 2015 during his lectures in several countries (for instance the USA, New Zealand, Kazakhstan…). Students were trading on the OANDA FX Trade Practice platform with currency pairs and CDFs. Initially the students were given

$100 000, meanwhile the trading period was standardized and took three months. As those students did not trade with real money, they were motivated to achieve as good result as possible by a financial reward together with extra points for the exam on account of the winner (student with the highest account balance at the end of the trading period). One of the learning objective was to experience the first-hand trading. At the end of the trading period students had to submit in detail recorded trading history together with their answers to a short questionnaire and demographic information. There were collected also the student's login information and passwords of his/her official game account, therefore it was not possible to change the account later, reset the losses, use more accounts etc. Based on the collected information there was created a unique dataset containing experimental trading data linked to the individual student traders. There were 292 students involved in the research over the time, who made in total 12 416 trades with the total volume of over three billion units. From the above-mentioned students, 43% (125 students) became profit makers. That is to say that their account balance at the end of trading period was higher than $100 000. Regarding the gender diversity of traders, there were 120 females and 172 males involved in the research. Overall, there were 29% of trades made by females, 71% by males.

The setting of the game in the currency markets is convenient; currency markets are liquid (eg. Mancini et al., 2013) and close to efficient. It is difficult to make meaningful price predictions and trading is more a matter of luck then skill. Accordingly the skill component does not distort the picture and the trading patterns and strategies tend to be more behavioural in nature.

The data generation and collection process proceeded as follows: The course was started with a series of lectures and assignments designed to explain the currency trading basics and the use of the trading platform. The game was launched sometime between the second and fourth week, and was running for the rest of the semester (60 to 90 days). Soon the focus shifted on the other topics and the game continued in the background. The rules and the interference with the students were minimal, students were not asked for any specific strategy, neither encouraged nor discouraged to the use of fundamental or technical analysis; there was no "desired" amount, frequency, size, or currency of trades. The winner was the student with the highest trading account balance at the end of the given time period. The ending profit or loss did not affect the course grade except that the winner (and only the winner) earned few extra points towards the final course grade and in some cases a voucher to a bookshop.

The experimental setting has obvious disadvantage, the money is not real and thus the joy of winning (pain of losing) is moderated. This should be slightly counteracted by awarding the winner. The counterargument might be that the "winner takes it all" reward scheme is problematic, there is no incentive for scoring second (third...); similarly, scoring low does not bear any penalty. This and the fact that it is hard to predict currency rates even for professional traders means that those students were in effect encouraged taking higher-than-normal risk and engaged in "all or nothing" gamble. It was not possible to rule those problems out, however, there is no indication of more frequent occurrence of large bets on the last few days of the game, which would point out a tendency towards pure gambling. It can be assumed that the students derived some benefits also from simply doing well, even if not the best. This could result from the long-term continuance of the experiment and the psychological benefit (cost) of favourable (unfavourable) comparison with the peers, and, perhaps most importantly, by keeping the current winner and his/her balance at strict confidentiality. This conjecture is supported by student comments and informal feedback. The assumption is that in spite of the singular incentive, the students chose investment strategies without trying to "game the system" or engaging in an ultimate all-in deciding gamble. The indirect evidence will be shown in subsequent chapters.

One can identify various advantages of the experimental setting. Firstly, participants do not self-select into roles and thus the sample is less biased: take, for example an effect of gender on trading decisions. To compare behaviour of actual female and male traders is problematic because the female traders might have some male characteristics, which made them to select the trading profession and helped them to succeed in it (self-selection and survivorship bias). The sample of students/traders does not represent the population ideally, because all of them decided to study business, but not all of them would like to do the trading career. Second advantage of the experimental setting is the homogeneity of objectives. In real situation, someone might set up a currency position as a hedge for some other asset. For example, if somebody's savings are predominantly in EUR but expected spending in USD, that person might want to open a short position in EUR/USD in order to hedge the Euro exposure in the savings account. The loss on currency position offsets the gain on the savings account (and vice versa), consequently decreasing the volatility of wealth. In case that the real traders behaviour is analysed, their goals are not known, are they speculating or hedging? Different objectives would lead to different trading strategies.

Moreover, in the real life, trader's wealth is given by the sum of different assets, hence a loss in Oanda may be compensated by gain in another asset and a high net asset value person may trade differently from a low net asset value person.

In spite of not being offered monetary compensation, the data shows that students took the trading game seriously. It is supposed that it was partly caused by the fact, that students found the game interesting, as shown by the questionnaire result, 75% of students answered that they traded because it was a course requirement and also interesting for them, 5% traded because it was purely interesting and 20% of students traded only because it was a course requirement. According to another question, 44% of students actually had at least some feelings of addiction during the game.

Students were strongly self-motivated partly because they played the game while being introduced to the world of international finance as the course progressed (thus seeing its relevance and connection with the real world). The other part of their motivation could have been the peer pride. Students were often discussing their trading success and failures or boastfully showing others their impressive results on their Oanda-enabled smart phones. The fact that they competed with and benchmarked their results against their classmates and friends over an extended period possibly made the game more interesting (compared to some short laboratory experiment with strangers).

2.2 TimeLtoP Ratio

For simplifying of calculations there was specified a special variable named timeLtoP, which is calculated as shown below in the formula:

$$timeLtoP = \frac{average\ holding\ period\ of\ unprofitable\ trades}{average\ holding\ period\ of\ profitable\ trades}. \quad (1)$$

The resulting ration provides information about presence of the disposition effect. In case of the result equal to 1, there is no disposition effect, the higher timeLtoP, the higher probability of the disposition effect presence. Additionally, the ration less than 2 shows the reverse disposition effect. For example, timeLtoP equal 2 can be intepreted as an average holding period of unprofitable trades two times longer than average holding period of profitable trades.

3 Results and Discussion- Disposition Effect with Regard to Gender

In this Chapter the difference in the disposition effect bias with respect to the gender of trader is examined. There are many studies focused on examination of the gender differences in the trading approach, for example, Jacobsen et al. (2014) compared the optimism of males and females in many fields including the economy and financial markets. According to his research males tend to be significantly more optimistic than women. Halko (2012), Dwyer et. al. (2002) and Barber and Odean (2000) confirmed that females are more risk averse than man.

First of all the two sample F-Test for testing the equality of variances on the significance level 0.05 was done with following hypothesis:

$$H_0: \sigma_m = \sigma_f,$$

$$H_1: \sigma_m \neq \sigma_f,$$

where σ_m represents variance of the trades made by males and σ_f represents variance of trade made by females.

	Females	Males
Mean	5.69	2.64
Variance	144.65	44.43
F	3.26	
P-value	0	
F_{crit}	1.05	

Table 1 F-Test – difference in gender

As shown in Table 1, $F>F_{crit}$, therefore it is possible to reject H_0. The test is statistically significant, as p-value<α. Based on this results the two sample t-test for different variances was chosen to test the difference in mean value of length of trades. The statistic hypotheses were set as follows:

$$H_0: \mu_m = \mu_f \rightarrow \textit{Gender has no effect on the length of trades,}$$

$$H_1: \mu_m \neq \mu_f \rightarrow \textit{Gender has an effect on the length of trades,}$$

where μ_m represents mean length of trades made by males and μ_f represents mean length of trades made by females. Those hypotheses were determined based on the assumption, is of length of trades within two groups are significantly different, those group have different tendency to succumb to the disposition effect bias.

T	13.28
P-value	0.00
T_{crit}	1.96

Table 2 T-Test – Difference in Gender

This t-test shown in Table 2 is statistically significant and based on this is is possible to reject H_0, as $T>T_{crit}$.

Table 3 also shows the length of trades for males and females and it is obvious that females tend to the disposition bias much more than males, they held loss trades more than two time longer than profitable trades. Contrary males held unprofitable trades only 1.3 times longer than those, which were profitable.

	males			females		
	N	**Length of trade (days)**	**Avg. size**	**N**	**Length of trade (days)**	**Avg. size**
Profit	3 222	2.5	$1M	1 177	5.8	$350K
Loss	1 588	3.2	$850K	556	12.3	$200K
timeLtoP		1.3			2.1	

Table 3 Holding Period of Trades According to the Gender

Regarding the Table 3, there is also the average size of the trades presented, while there is visible significant difference between different gender approaches. The average size of profitable trades of males is 1M$, which is almost three times more than the average size of females´ profitable trades in height of 350k$. The average size of loss trades of males is 850k$, contrary of females is 200k$. This fact leads to the presumption that females are much more risk averse than males. As students had to fill in a questionnaire, which involved also one question regarding to the risk attitude, it is possible to compare their answers to the real numbers. Following Figures 9 and 10 show answers to the question, whether they enjoy taking risk.

Figure 1 Risk approach answers-females

Figure 2 Risk approach answers-males

Those answers were statistically tested. Firstly the two sample F-test for testing the difference in variances of two groups-males and females was used. According to this test, which was found statistically significant, $F<F_{crit}$ so it is not possible to reject the H_0 hypothesis, representing the variance equality. Subsequently the ANOVA was chosen to test the difference of means. As well this test was statistically significant and the $F>F_{crit}$. Based on this results it is possible to make the conclusion that difference in answers on this question are statistically significant.

According to the result of the questionnaire, males involved in the research tend to behave more risky than females, as is shown in the Figure 1 and Figure 2. From the total number of respondents, 46 % of females stated that they do not enjoy taking risk, the same answer gave only 26 % of males. On the other hand, 41% of females confirmed enjoying taking risk, the same answered 60 % of males. This result is consistent with the Halko (2012), Dwyer et. al. (2002) and Barber and Odean (2000), who confirmed that females are more risk averse than males. The subjective approach to the risk goes also together with the Jacobsen et al. (2014) and the optimistic approach of males. Answers collected from students were also statistically tested. Based on the F-test for variance it is possible to say that variances of both samples (for males and for females) are identical, as we could not reject the H_0 hypothesis. According to the test ANOVA the differences within genders are statistically significant on the significance level 0,05

Looking at the results of the questionnaire and results of trading, they are in accordance, as it is obvious that males take much bigger trades and are less afraid of loss than females.

4 Conclusion

Kaheman and Tversky (1985), within the prospect theory, proposed for the first time the hypothesis that investors tend to accept more risk in case of facing losses than they would ever accept in case of gain. This premise laid the ground for several researches focused on a so-called disposition effect. This effect causes that investors tend to sell winning investments too soon, because they are risk averse and afraid of fall, and contrary selling loosing investments too late, because they are willing to take the risk of losing even more in exchange for the chance of its increase in the close future.

This paper was focused on the difference in disposition effect in trading with respect to gender. Data of males and females were separately tested on the disposition effect presence per TimeLtoP ratio. According to our results there is significant difference in the length of profitable and unprofitable trades, as well as the risk approach of individual gender.

Acknowledgement

This work has been supported by the Czech Science Foundation (GACR) under project 17-19981S and SP2018/34, an SGS research project of VSB-TU Ostrava.

References

[1] Barberb, B.M. and Odean T. (2000). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 55, 261–292.

[2] Barberis, N. and Xiong, W. (2009), What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation. *The Journal of Finance*, 64, 751–784.

[3] Chen, G., Kim, K. A., Nofsinger, J. R. and Rui, O. M. (2007) Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making*, 20, 425–451.
[4] Choe, H. and Eom, Y. (2009), The disposition effect and investment performance in the futures market. *J. Fut. Mark.*, 29, 496–522.
[5] Dwyera, P. D, Gilkeson J.H. and J.A. List (2002). Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters*, 76, 151–158.
[6] Dvořáčková, H., Jochec, M., Tichý, T. (2019). Disposition Effect in Currency Trading: An Evidence from Experimental Student Games. *Review of Research and Social Intervention*, 64, 246-261.
[7] Dvořáčková H. and M. Jochec (2018). Disposition Effect in the Experimental Currency Trading. *In: 9th International Scientific Conference Managing and Modelling of Financial Risks.* VSB-Technical University of Ostrava, 2018.
[8] Field, A.P. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll* 4th ed., London: SAGE.
[9] Fogel, S. O. and T. Berry (2006). The Disposition Effect and Individual Investor Decisions: The Roles of Regret and Counterfactual Alternatives. *Journal of Behavioral Finance*, 7, 107-116.
[10] Halko M.L. et al. (2012). The gender effect in risky asset holding. *Journal of Economic Behavior & Organisation*, 83, 66–81.
[11] Jacobsen B. et al. (2014). Gender differences in optimism and asset allocation. *Journal of Economic Behavior & Organisation*, 107, 630–651.
[12] Kahneman, Daniel, and Amos Tversky (1979) Prospect theory: An analysis of decision under risk, *Econometrica*, 46, 171–185.
[13] Kaustia, M. (2010) *Disposition Effect, in Behavioral Finance: Investors, Corporations, and Markets* (eds H. K. Baker and J. R. Nofsinger), John Wiley & Sons, Inc., Hoboken, NJ, USA.
[14] Landau, I. (2006). *Ways of being rude*, Technical report. Ben Gurion University, Israel.
[15] Lehenkari, M. (2007) *The Disposition Effect: Underlying Mechanisms and Implications for Individual Investors*. Working paper. University of Oulu.
[16] Mancini et al. (2013) Liquidity in the foreign exchange market: measurement, commonality, and risk premiums. *Journal of Finance*, 68, 1805-1841.
[17] Marciukaityte, D. and Szewczyk, S. H. (2012), Disposition Effect and Public Secondary Equity Offerings. *Journal of Financial Research*, 35, 55–77.
[18] Neubauer, J., Sedlačík, M. & Kříž, O. (2012). *Základy statistiky: aplikace v technických a ekonomických oborech*, Praha: Grada Publishing.
[19] OANDA. [Online]. [cit. 26.11.2018]. Available at: https://www.oanda.com.
[20] Odean, Terrance, 1998, Are investors reluctant to realize their losses, *Journal of Finance*, 53, 1775–1798.
[21] Shefrin, H., and M. Statman (1985) The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance*, 40, 777–790.
[22] Shefrin H. (2008) *A Behavioral Approach to Asset Pricing (Second Edition)*, Academic Press.
[23] Tavares, H. et al. (2001). Gender differences in gambling progression. *Journal of Gambling*, 17, 151-159.

Author Index