Mathematical Methods in Economics

MME 2023

Proceedings of the

41st International Conference on

Mathematical Methods in Economics

Czech Society for Operations Research

Prague University of Economics and Business

Proceedings of the

41st International Conference on Mathematical Methods in Economics

September 13–15, 2023

Prague, Czech Republic

41st International Conference on Mathematical Methods in Economics

September 13–15, 2023

Hosted by the Prague University of Economics and Business, Faculty of Informatics and Statistics

Winston Churchill Square 1938/4, 130 67 Prague 3, Czechia

mme2023.vse.cz

Mathematical Methods in Economics (MME 2023)

Proceedings of the 41st International Conference on Mathematical Methods in Economics

Published by the Czech Society for Operations Research

Winston Churchill Square 1938/4, 130 67 Prague 3, Czechia

www.csov.eu

Edited by Jana Sekničková and Vladimír Holý

Prague 2023

This publication is not subject to a language check.

All papers have passed a blind peer review process.

All papers adhere to the ethical guidelines of the publisher.

ISBN 978-80-11-04132-8

ISSN 2788-3965

Programme Committee

- prof. RNDr. Helena Brožová, CSc.
  Czech University of Life Science Prague, Faculty of Economics and Management
- prof. Ing. Mgr. Martin Dlouhý, Dr., MSc.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- doc. Ing. Jan Fábry, Ph.D.
  ŠKODA AUTO University
- prof. RNDr. Ing. Petr Fiala, CSc., MBA
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- prof. Ing. Jana Hančlová, CSc.
  Technical University of Ostrava, Faculty of Economics
- Mgr. Vladimír Holý, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- prof. Ing. Josef Jablonský, CSc.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- doc. RNDr. Jana Klicnarová, Ph.D.
  University of South Bohemia, Faculty of Economics
- Ing. František Koblasa, Ph.D.
  Technical University of Liberec, Faculty of Mechanical Engineering
- doc. RNDr. Ing. Miloš Kopa, Ph.D.
  Charles University, Faculty of Mathematics and Physics
- Ing. Martina Kuncová, Ph.D.
  College of Polytechnics Jihlava
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- prof. RNDr. Jan Pelikán, CSc.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- prof. Dr. Ing. Miroslav Plevný
  University of West Bohemia, Faculty of Economics
- prof. RNDr. Jaroslav Ramík, CSc.
  Silesian University in Opava, School of Business Administration in Karviná
- Mgr. Jana Sekničková, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics

- Ing. Karel Sladký, CSc.
  Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation
- Ing. Ondřej Sokol, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- doc. Ing. Tomáš Šubrt, Ph.D.
  Czech University of Life Science Prague, Faculty of Economics and Management
- doc. RNDr. Jana Talašová, CSc.
  Palacký University in Olomouc, Faculty of Science
- Ing. Miroslav Vavroušek, Ph.D.
  Technical University of Liberec, Faculty of Mechanical Engineering
- prof. RNDr. Milan Vlach, DrSc.
  Charles University in Prague, Faculty of Mathematics and Physics
  The Kyoto College of Graduate Studies for Informatics
- prof. RNDr. Karel Zimmermann, DrSc.
  Charles University in Prague, Faculty of Mathematics and Physics
- prof. Ing. Miroslav Žižka, Ph.D.
  Technical University of Liberec, Faculty of Economics

Organizing Committee

- prof. Ing. Mgr. Martin Dlouhý, Dr., MSc.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- Mgr. Vladimír Holý, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- prof. Ing. Josef Jablonský, CSc.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- Ing. Martina Kuncová, Ph.D.
  College of Polytechnics Jihlava
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- Mgr. Jana Sekničková, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- Ing. Ondřej Sokol, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics
- Ing. Bc. Petra Zýková, Ph.D.
  Prague University of Economics and Business, Faculty of Informatics and Statistics

Ethical Guidelines

The Czech Society for Operations Research as a publisher of the Proceedings of the International Conference on Mathematical Methods in Economics is committed to the highest ethical standards. All authors, reviewers, and editors must comply with the following ethical principles. In case of any doubts regarding the Ethical Guidelines, do not hesitate to contact the editors of the proceedings.

Ethical Expectations

Editors' Responsibilities

- To acknowledge receipt of submitted manuscripts within two working days and to ensure an efficient, fair, and timely review process.
- To identify manuscripts that are fully within the scope and aim of the conference. To treat all submissions fairly without any favour of prejudice regarding authors' gender, religious or political beliefs, ethnic or geographical origin.
- To only authorise the review and publication content of the highest quality.
- To recuse himself or herself from processing any manuscript if he or she has any conflict of interest with any of the authors or institutions related to the manuscripts.
- To provide advice to the authors during the submission process when necessary.
- To be transparent with regards to the review and publication process with an appropriate care that individuals will not be identified when it is inappropriate to do so.
- To not use any parts or data of the submitted manuscript for his or her own future research as the submitted manuscript is not published yet.
- To respond immediately and take reasonable action when ethical problems occur concerning a submitted or published manuscript.

Authors' Responsibilities

- To carefully read the Instructions for authors published on the website of the conference.
- To claim that the submitted manuscript is not under consideration or accepted for publication elsewhere. To acknowledge and cite accordingly all content overlaps with already published content.
- To ensure that the submitted manuscript is original, prepared to a high scholarly standard and fully referenced using the prescribed referencing convention.
- To obtain permission to reproduce any content from other published sources. To clarify that all data used in the manuscript has been acquired following ethical research standards.
- To accurately acknowledge funding sources related to the submitted manuscript.

- To carefully read all the conditions included in the copyright form and to accept the copyright form during the submission process.
- To declare any potential conflicts of interest (real or apparent) at any stage during the publication process.
- To recognise that the editor has the final decision to publish the submitted, reviewed and accepted manuscript.
- To immediately inform the editors whenever any obvious error in his or her published manuscript is identified. To cooperate with the editors in the retraction or correction of the manuscript if necessary.

Reviewers' Responsibilities

- To immediately notify the editors whenever the reviewer feels unqualified to review the assigned manuscript or sees difficulties in meeting the deadline for completion of the review.
- To agree to review a reasonable number of manuscripts at the same time.
- To inform the editors if there is any possible conflict of interest related to the assigned manuscript. Specifically, to avoid reviewing any manuscript authored or co-authored by a person with whom the reviewer has an obvious personal or academic relationship.
- To treat the manuscript in a confidential manner and to not use any parts or data of the reviewed manuscript for his or her own future research as the reviewing manuscript is not published yet.
- To assist in improving the quality of the manuscript by reading the manuscript with appropriate care and attention, reviewing the manuscript objectively and being constructively critical.
- To immediately notify the editors of any similarities between the reviewing manuscript and another article either published or under consideration by another journal.

Procedures for Dealing with Unethical Behaviour

Dealing with Possible Misconduct

- Editors have a duty to act if they suspect any misconduct or if a claim of misconduct has been reported by anyone. This duty applies to both published and unpublished articles.
- Editors should not simply reject an article that raises concerns about possible misconduct. Editors are ethically obliged to pursue alleged cases.
- Whoever informs the editors of such conduct should provide sufficient information and evidence to initiate an investigation.
- Editors must take all allegations seriously and treat them similarly until a successful decision or conclusion is reached.
- Editors should first seek a response from those suspected of misconduct. If the editors are not satisfied with the response, they should ask the relevant employer, institution, or some appropriate body to investigate.
- Editors should make all reasonable efforts to ensure that a proper investigation into alleged misconduct is conducted. If this does not happen, editors should make all reasonable attempts to persist in obtaining a resolution to the problem.

Retraction of an Article

- Editors have clear evidence that the findings are unreliable, either as a result of misconduct or honest error.
- The findings have previously been published elsewhere without proper cross-referencing, permission or justification (i.e. redundant publication). The journal that first published the article may issue a notice of redundant publication, but should not retract the article unless the findings are unreliable. Any journals that subsequently publish a redundant article should retract it and state the reason for the retraction.
- The article constitutes plagiarism and reports unethical research.
- The article should be retracted as soon as possible after the editors are convinced that the article is seriously flawed and misleading.

Outcomes of Unethical Behaviour in Increasing Order of Severity

- Informing or educating the author or reviewer where appears to be a misunderstanding or misapplication of acceptable standards.
- A more strongly worded letter to the author or reviewer covering the misconduct and as a warning to future behaviour.
- Retraction or withdrawal of a publication from the proceedings, in conjunction with informing the head of the author or reviewer's department, as well as Abstracting & Indexing services.
- Imposition of a formal embargo on contributions from an individual in the proceedings for a defined period.

Foreword

We are delighted to present to you the Conference Proceedings, which have collected the best selected contributions presented at the 41st International Conference on Mathematical Methods in Economics organized by the Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, under the auspices of the Czech Society for Operations Research, the Slovak Society for Operations Research, and the Czech Econometric Society. The conference was held in Prague from September 13 to 15, 2023.

This traditional meeting brings together academicians and professionals interested in the theory and applications of operations research and econometrics. It serves as a significant event in the field. This year, we welcomed more than 110 researchers from 8 countries who also served as discussants of the papers and helped to improve the quality of the research results presented during the conference days. The contributions followed new trends in econometrics and operations research and built bridges between researchers, academicians, and practitioners in the industrial and institutional sectors, sharing recent theoretical and applied results. The wide range of topics presented at the conference demonstrates the importance of using mathematical methods in many fields of economics.

We hosted two distinguished plenary speakers who contributed to the conference program with their speeches. Prof. Dimitris K. Despotis (University of Piraeus, Greece) gave a talk on Network Data Envelopment Analysis with a focus on the prevalent methodological approaches and some recent developments. Prof. Miloš Kopa (Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic) presented recent research on Decision Making in Finance via Stochastic Dominance. Moreover, special attention was paid to two sections collecting the best student talks submitted to the PhD student competition organized by the Czech Society for Operations Research. We congratulate the winner – Mgr. Jana Junová (Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic), and the other laureates.

Finally, we would like to express our deep thanks to the organizers, all the reviewers, and the members of the scientific committee for their contribution to the successful organization of this high-level scientific conference.

Prague, September 2023 Miloš Kopa

Contents

A Consensual Coherent Priority Vector of Pairwise Comparison Matrices in Group Decision-Making

David Bartl[1], Jaroslav Ramík[2]

Abstract. The Analytic Hierarchy Process (AHP) is a method proposed to solve complex multi-criteria decision-making problems. Pairwise comparison methods are often used in AHP to derive the priorities of the successors of an element in the hierarchy. In this paper, we are concerned with group decision-making; that is, given n objects, such as criteria and/or variants, let m decision makers evaluate the n objects (pairwise) with respect to a criterion. The task is then to find a consensual priority vector of the m given $n \times n$ reciprocal pairwise comparison matrices. Recalling several desirable properties of the priority vector – consistency, intensity, and coherence – we consider the weakest one of the three, i.e. coherence, in the rest of the paper. In other words, given m coherent priority vectors, each provided by a decision maker of the group, the purpose is to find a single consensual priority vector of the group. To cope with this task, we propose a grade to measure the consensuality of a priority vector. We thus obtain an optimization problem, whose solution yields an optimal consensual ranking of the n given objects.

Keywords: multi-criteria group decision-making, pairwise comparison matrices, consensual priority vector, coherence, Analytic Hierarchy Process (AHP)

JEL Classification: C44, C65, C63, D79
AMS Classification: 90C29, 90C70

1 Introduction

The Analytic Hierarchy Process (AHP) is a popular and powerful tool to solve multi-criteria decision-making problems [9]. We consider the following main subproblem of the AHP, which is to be solved in every internal node of the hierarchy; that is, a node having some subnodes. Let n denote the number of these subnodes, which correspond to n objects $c_1, c_2, \dots, c_n$, i.e. criteria, subcriteria, and/or alternatives (variants). Notice that the internal node corresponds to some criterion, subcriterion, and/or the goal of the hierarchy. Henceforth, we shall use the single term criterion for simplicity. Given the information on the relative importance of the two items in each pair of the objects with respect to the given criterion (subcriterion, and/or the goal) in the form of an $n \times n$ pairwise comparison matrix A, the purpose is to calculate the priority vector, which is a vector of n weights $v_1, v_2, \dots, v_n$ assigned to the n objects $c_1, c_2, \dots, c_n$, respectively. The prominent methods to calculate the priority vector include Saaty's Eigenvector Method (EVM) and the Geometric Mean Method (GMM), see [9] and [8]. The priority vector provided by these methods, however, usually do not satisfy desirable properties – consistency, intensity, and/or coherence, in particular – see [10], [5], and [1].

For $i, j = 1, 2, \dots, n$, let a_{ij} be a value (i.e. quantity or number) that represents the decision maker's opinion how many times object c_i is more important or better than object c_j with respect to the given criterion. We thus obtain a (crisp) *pairwise comparison matrix* $A = \{a_{ij}\}$. This is a special case of the fuzzy case studied in [2], where the authors have proposed a new algorithm for computing priority vectors, satisfying desirable properties, of a fuzzy pairwise comparison matrix. In [3], the authors have improved and extended their new algorithm to the case when there are m decision makers (evaluators), and each of them assesses the relative importance of the two items in each pair of the objects with respect to the given criterion. In other words, given $n \times n$ pairwise comparison matrices $A^1, A^2, \dots, A^m$ such that the element a_{ij}^k of the k-th matrix represents the k-th decision maker's opinion how many times c_i is more important or better than c_j with respect to the given criterion for $i, j = 1, 2, \dots, n$ and for $k = 1, 2, \dots, m$, the extended algorithm provides a *joint* priority vector, satisfying the desirable properties, of the m pairwise comparison matrices $A^1, A^2, \dots, A^m$. In this paper, it is our purpose to provide an algorithm to compute a *consensual* priority vector, satisfying the desirable property of coherence only.

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czechia, bartl@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czechia, ramik@opf.slu.cz.

2 Abelian linearly ordered groups

In order to unify and generalize various approaches known from the literature, we use the elements of an Abelian linearly ordered group to evaluate the relative importance of the two items in each pair of the objects with respect to the given criterion, see [4] and [7]. Recall that an Abelian group is a pair $(G, \odot)$ where G is a non-empty set and $\odot$ is a commutative and associative binary operation on G satisfying also the existence of the neutral element $e \in G$ and the existence of the inverse element $a^{(-1)} \in G$ for each $a \in G$. We then have $a \odot e = a$ and $a \odot a^{(-1)} = e$ for every $a \in G$. We also put $a \div b = a \odot b^{(-1)}$ for all $a, b \in G$. An Abelian linearly ordered group (alo-group) is a triple $(G, \odot, \leq)$ such that $(G, \odot)$ is an Abelian group and $\leq$ is a binary relation of linear ordering on G such that $a \leq b$ implies $a \odot c \leq b \odot c$ for all $a, b, c \in G$. The well-known examples of alo-groups are the Multiplicative alo-group $\mathcal{R}_+ = (\mathbb{R}_+, \cdot, \leq)$ with the usual multiplication and the neutral element $e = 1$, the Additive alo-group $\mathcal{R} = (\mathbb{R}, +, \leq)$ with the usual addition and the neutral element $e = 0$, and the Fuzzy Multiplicative alo-group $\mathcal{F}_{]0;1[} = (]0;1[, \odot, \leq)$ with $a \odot b = ab/(ab + (1-a)(1-b))$ for $a, b \in]0;1[$ and the neutral element $e = \frac{1}{2}$, see [4], [7], and [8].

3 Desirable properties of the priority vector

Let us consider an alo-group $\mathcal{G} = (G, \odot, \leq)$ and let us denote the set of the first n positive natural numbers by $\mathcal{N}$; that is, we put $\mathcal{N} = \{1, 2, \dots, n\}$. Considering the set $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$, let $A = \{a_{ij}\}$ be an $n \times n$ matrix such that each of its element $a_{ij} \in G$ evaluates the relative importance of the objects c_i and c_j with respect to the given criterion. The matrix $A = \{a_{ij}\}$ is called a *pairwise comparison matrix*, or *PC matrix* for short, if it is *reciprocal*; that is, if the following two conditions hold for each $i, j \in \mathcal{N}$:

$$a_{ii} = e, \qquad \text{and} \qquad a_{ij} \odot a_{ji} = e. \tag{1}$$

Then the result of a pairwise comparison method based on the PC matrix $A = \{a_{ij}\}$ is a vector $v = (v_1, v_2, \dots, v_n)$ of the weights $v_1, v_2, \dots, v_n \in G$ of the objects $c_1, c_2, \dots, c_n \in \mathcal{C}$, respectively. In other words, the i-th component v_i of the priority vector v is the weight of the object c_i for $i \in \mathcal{N}$. We say the priority vector $v = (v_1, v_2, \dots, v_n)$ is *normalized* if $\bigodot_{i=1}^{n} v_i = e$.

Based upon the ideas that have already appeared in the literature ([10], [1], [5], [6], [2] and [3]), we define the notions of desirable properties as follows.

Definition 1. Let $A = \{a_{ij}\}$ be a PC matrix on an alo-group $\mathcal{G} = (G, \odot, \leq)$ and let $v = (v_1, v_2, \dots, v_n)$, with $v_j \in G$, be a priority vector.

(i) We say that the vector v is a *consistent vector* (CsV) of the PC matrix A if the following condition holds:

$$a_{ij} = v_i \div v_j \qquad \text{for all} \quad i, j \in \mathcal{N}. \tag{2}$$

(ii) We say that the vector v is an *intensity vector* (InV) of the PC matrix A if the following condition holds:

$$a_{ij} > a_{kl} \quad \text{if and only if} \quad v_i \div v_j > v_k \div v_l \qquad \text{for all} \quad i, j, k, l \in \mathcal{N}. \tag{3}$$

(iii) We say that the vector v is a *coherent vector* (CoV) of the PC matrix A if the following condition holds:

$$a_{ij} > e \quad \text{if and only if} \quad v_i > v_j \qquad \text{for all} \quad i, j \in \mathcal{N}. \tag{4}$$

If there exists a consistent, intensity, or coherent vector of the PC matrix A, then A is called a *consistent*, *intensity*, or *coherent PC matrix*, respectively.

By reciprocity (1) and by Definition 1, the following result is easy to see. This is why we omit its proof.

Proposition 2. *Let $A = \{a_{ij}\}$ be a PC matrix on an alo-group $\mathcal{G} = (G, \odot, \leq)$ and let $v = (v_1, v_2, \dots, v_n)$, with $v_j \in G$, be a vector. Then:*

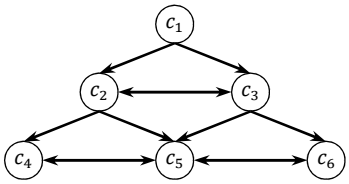
(i) *If v is a consistent priority vector of the PC matrix A, then it is an intensity priority vector of A.*
(ii) *If v is an intensity priority vector of the PC matrix A, then it is a coherent priority vector of A.*

A coherent matrix of pairwise comparisons of elements $c_1, \dots, c_6$ with respect to the criterion by the 1st expert:

$$A_1 = \begin{pmatrix} 1 & 5 & 5 & 5 & 5 & 5 \\ 1/5 & 1 & 1 & 5 & 5 & 5 \\ 1/5 & 1 & 1 & 5 & 5 & 5 \\ 1/5 & 1/5 & 1/5 & 1 & 1 & 1 \\ 1/5 & 1/5 & 1/5 & 1 & 1 & 1 \\ 1/5 & 1/5 & 1/5 & 1 & 1 & 1 \end{pmatrix}$$

The induced quasi-linear ordering:

$$\underbrace{c_1}_{\mathcal{K}_1} \succ \underbrace{c_2 \approx c_3}_{\mathcal{K}_2} \succ \underbrace{c_4 \approx c_5 \approx c_6}_{\mathcal{K}_3}$$



The matrix P_1 representing the induced quasi-linear ordering:

$$P_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

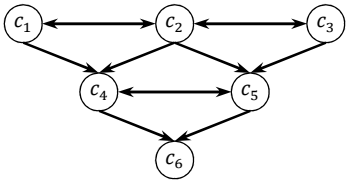
Figure 1 The 1st expert's judgements of 6 elements $c_1, c_2, c_3, c_4, c_5, c_6$ with respect to some criterion

A coherent matrix of pairwise comparisons of elements $c_1, \dots, c_6$ with respect to the criterion by the 2nd expert:

$$A_2 = \begin{pmatrix} 1 & 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 1 & 5 & 5 & 5 \\ 1 & 1 & 1 & 5 & 5 & 5 \\ 1/5 & 1/5 & 1/5 & 1 & 1 & 5 \\ 1/5 & 1/5 & 1/5 & 1 & 1 & 5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 1 \end{pmatrix}$$

The induced quasi-linear ordering:

$$\underbrace{c_1 \approx c_2 \approx c_3}_{\mathcal{K}_1} \succ \underbrace{c_4 \approx c_5}_{\mathcal{K}_2} \succ \underbrace{c_6}_{\mathcal{K}_3}$$



The matrix P_2 representing the induced quasi-linear ordering:

$$P_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 2 The 2nd expert's judgements of 6 elements $c_1, c_2, c_3, c_4, c_5, c_6$ with respect to some criterion

4 An algorithm to generate a consensual priority vector of coherent PC matrices in group decision-making

Let $\mathcal{G} = (G, \odot, \leq)$ be an alo-group. Given an $n \times n$ pairwise comparison matrix $A = \{a_{ij}\}$ on the alo-group $\mathcal{G}$, we introduce binary relations $\succ$, $\approx$ and $\succcurlyeq$ on the set $\mathcal{N}$ as follows. For $i, j \in \mathcal{N}$, we define that $i \succ j$ if and only if $a_{ij} > e$, and we define that $i \approx j$ if and only if $a_{ij} = e$, where e is the neutral element of the alo-group $\mathcal{G}$. Finally, for $i, j \in \mathcal{N}$, we define that $i \succcurlyeq j$ if and only if $i \succ j$ or $i \approx j$. Notice that the PC matrix A is coherent if and only if the relation $\succcurlyeq$ is a quasi-linear ordering of the set $\mathcal{N}$, and also $\approx$ is a relation of equivalence on $\mathcal{N}$; that is, the relation $\succcurlyeq$ is complete ($i \succcurlyeq j$ or $j \succcurlyeq i$) and transitive ($i \succcurlyeq j \succcurlyeq k \implies i \succcurlyeq k$), and the relation $\approx$ is reflexive ($i \approx i$), symmetric ($i \approx j \implies j \approx i$), and transitive ($i \approx j \approx k \implies i \approx k$).

Assume in the sequel that the PC matrix $A = \{a_{ij}\}$ is coherent. Then the relation $\succcurlyeq$ of quasi-ordering of the set $\mathcal{N}$ can equivalently be represented by an $n \times n$ matrix $P = \{p_{ij}\}$ consisting of 0's and 1's as follows. Let $\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_R$ be all the pairwise distinct equivalence classes of the relation $\approx$ on the set $\mathcal{N}$, and let the classes be ordered so that $r < s$ implies $k \succ l$ for all $k \in \mathcal{K}_r$, for all $l \in \mathcal{K}_s$, and for all $r, s = 1, 2, \dots, R$. Let $i_1 := n$, and for $r = 1, 2, \dots, R$, let $i_{r+1} := i_r - |\mathcal{K}_r|$, where $|\mathcal{K}_r|$ denotes the number of the elements of the class $\mathcal{K}_r$. Notice that $i_{R+1} = 0$. Then, for $r = 1, 2, \dots, R$, we put $p_{i_r j} = 1$ for $j \in \mathcal{K}_r$, and we put $p_{i_r j} = 0$ for $j \in \mathcal{N} \setminus \mathcal{K}_r$. Finally, we put $p_{ij} = 0$ for all $i \in \mathcal{N} \setminus \{i_1, i_2, \dots, i_R\}$ and for all $j = 1, 2, \dots, n$. Notice that P is a binary matrix, which consists of elements 0 and 1, and satisfies the following system of inequalities and conditions:

$$\sum_{i=1}^{n} p_{ij} = 1 \qquad \text{for} \quad j \in \mathcal{N}, \tag{5}$$

$$\sum_{j=1}^{n} p_{ij} \leq i \times \max\{0,\ 1 + k - \sum_{j=1}^{n} p_{i+k,j}\} \qquad \text{for} \quad k = 1, \dots, n-i \qquad \text{for} \quad i \in \mathcal{N}, \tag{6}$$

$$p_{ij} \in \{0, 1\} \qquad \text{for} \quad i, j \in \mathcal{N}. \tag{7}$$

In words, the maximal elements are in the class $\mathcal{K}_1$ and the smaller elements are in the subsequent classes $\mathcal{K}_2, \dots \mathcal{K}_R$. The elements of the class $\mathcal{K}_r$ are represented in the i_r-th row of the matrix P. Moreover, there being $|\mathcal{K}_r|$ elements in the class $\mathcal{K}_r$, therefore $|\mathcal{K}_r|$ 1's in the i_r-th row, then the subsequent $(|\mathcal{K}_r| - 1)$ rows, i.e. rows $i_r - 1, \dots, i_r - |\mathcal{K}_r| + 1$, of the matrix P must be zero. Examples presented in Figures 1 and 2 illustrate this procedure.

In this paper, it is our purpose to consider the main subproblem of the AHP extended as follows. Given the alo-group $\mathcal{G} = (G, \odot, \leq)$ and the n objects $c_1, c_2, \dots, c_n$ to be judged with respect to the given criterion by m independent decision makers (evaluators), each of the decision makers assesses the relative importance of the two items in each pair of the objects with respect to the given criterion by using an element of the alo-group $\mathcal{G}$; that is, let $a_{ij}^k \in G$ represent the k-th decision maker's opinion how many times c_i is more important or better than c_j with respect to the given criterion for $i, j = 1, 2, \dots, n$ and for $k = 1, 2, \dots, m$. Additionally, we assume that each of the

m decision makers is coherent; that is, let each PC matrix $A^k = \{a_{ij}^k\}$ be coherent and let $w^k \in G^n$ be a coherent priority vector of the PC matrix A^k for $k = 1, 2, \ldots, m$. Now, given the coherent PC matrices $A^1, A^2, \ldots, A^m$, or their coherent priority vectors $w^1, w^2, \ldots, w^m \in G^n$, our purpose is to find a single *consensual* priority vector $v = (v_1, v_2, \ldots, v_n) \in G^n$; that is, a priority vector $v \in G^n$ that is consensual with each of the priority vectors $w^1, w^2, \ldots, w^m$ as much as possible.

Generally speaking, we define that two priority vectors $w^k \in G^n$, and $v \in G^n$ are *consensual* if it holds $w_i^k > w_j^k \Leftrightarrow v_i > v_j$ for every $i, j = 1, 2, \ldots, n$. Actually, as all the PC matrices $A^1, A^2, \ldots, A^m$ are coherent, they induce the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ of the set $\mathcal{N}$, defined in the above given way. Therefore, we define that quasi-linear orderings $\succcurlyeq^k$ and $\succcurlyeq$ of $\mathcal{N}$ are *consensual* if it holds $i \succ^k j \Leftrightarrow i > j$ for every $i, j \in \mathcal{N}$. Recalling that $w^k \in G^n$ is a coherent priority vector of the PC matrix A^k, observe that $i \succ^k j$ implies $w_i^k > w_j^k$ for $i, j \in \mathcal{N}$ for $k = 1, 2, \ldots, m$. Consequently, it is possible to simplify our task as follows: given the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ of the set $\mathcal{N}$, our purpose is to find a single *consensual* quasi-linear ordering $\succcurlyeq$ of the set $\mathcal{N}$; that is, a quasi-linear ordering $\succcurlyeq$ of $\mathcal{N}$ that is consensual with each of the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ as much as possible.

Given quasi-linear orderings $\succcurlyeq^k$ and $\succcurlyeq$ of the set $\mathcal{N}$, we define the grade of their non-consensuality as follows: Let $P^k = \{p_{ij}^k\}$ and $P = \{p_{ij}\}$ be the binary matrix representing the relation $\succcurlyeq^k$ and $\succcurlyeq$, respectively, defined in the above given way; notice that the matrix P satisfies relations (5)–(7). We then define the grade of their non-consensuality as

$$\delta(P^k, P) = \sum_{j=1}^{n} \sum_{\substack{i^k=1 \\ p_{i^k j}^k = 1}}^{n} \sum_{\substack{i=1 \\ p_{ij}=1}}^{n} |i^k - i|\,, \tag{8}$$

where $|i^k - i|$ denotes the absolute value of the difference $i^k - i$. (Alternatively, we could replace $|i^k - i|$ by its square $|i^k - i|^2$, cube $|i^k - i|^3$, or any other power of it.) The idea behind (8) is to penalize the change of the "level" of the element $j = 1, 2, \ldots, n$ when transiting from one quasi-ordering (e.g. $\succcurlyeq^k$) to the other (e.g. $\succcurlyeq$). Then the total non-consensuality of the quasi-linear ordering $\succcurlyeq$ of $\mathcal{N}$ with the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ is defined as $\delta(P^1, P^2, \ldots, P^m, P) = \sum_{k=1}^{m} \delta(P^k, P)$; that is,

$$\delta(P^1, P^2, \ldots, P^m, P) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \sum_{i^k=1}^{n} |i^k - i| \times p_{i^k j}^k \right) \times p_{ij}\,. \tag{9}$$

To meet our purpose; that is, to find a single quasi-linear ordering $\succcurlyeq$ of the set $\mathcal{N}$ that is consensual with each of the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ as much as possible, we minimize (9) subject to (5)–(7).

We notice that the aforegiven optimization problem (minimize (9) subject to (5)–(7)) is integer and non-smooth, hence difficult to solve. For this reason, we may restrict $\succcurlyeq$ to be a consensual *linear* ordering of the set $\mathcal{N}$, so that the corresponding matrix P reduces to a simple *permutation* matrix. Constraints (6) then reduce to $\sum_{j=1}^{n} p_{ij} \leq i$ for $i \in \mathcal{N}$, which, by taking (5) and (7) into account, can further be simplified to $\sum_{j=1}^{n} p_{ij} = 1$ for $i \in \mathcal{N}$. Then, to find a single linear ordering $\succcurlyeq$ of the set $\mathcal{N}$ that is consensual with each of the quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \ldots, \succcurlyeq^m$ as much as possible, we minimize (9) subject to

$$\sum_{i=1}^{n} p_{ij} = 1 \qquad \text{for} \quad j \in \mathcal{N}, \tag{10}$$

$$\sum_{j=1}^{n} p_{ij} = 1 \qquad \text{for} \quad i \in \mathcal{N}, \tag{11}$$

$$p_{ij} \in \{0, 1\} \qquad \text{for} \quad i, j \in \mathcal{N}, \tag{12}$$

which is an assignment problem actually. It is well-known that the matrix of the coefficients by the variables p_{ij} in (10) and (11) is totally unimodular, so that (12) can be relaxed to

$$0 \leq p_{ij} \leq 1 \qquad \text{for} \quad i, j \in \mathcal{N}, \tag{13}$$

yet the easy problem of continuous linear programming (minimize (9) subject to (10), (11), and (13)) has an *integer* optimal solution.

Once we find an optimal solution to the above optimization problem (minimize (9) subject to either (5)–(7), or (10), (11), and (13)), we construct the corresponding consensual quasi-linear or linear ordering $\succcurlyeq$ of the set $\mathcal{N}$ as follows. Let $i_1, i_2, \ldots, i_R$ be all the pairwise distinct elements of the set $\{\, i \in \mathcal{N} \mid p_{ij} = 1 \text{ for some } j \in \mathcal{N} \,\}$ and let them be ordered so that $i_1 > i_2 > \cdots > i_R$. For $r = 1, 2, \ldots, R$, put $\mathcal{K}_r = \{\, j \in \mathcal{N} \mid p_{i_r j} = 1 \,\}$. Finally, let $i \approx j$

An optimal solution to the problem:

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

The quasi-linear ordering represented by the optimal solution P:

$$\underbrace{c_1 \approx c_2 \approx c_3}_{\mathcal{K}_1} \succ \underbrace{c_4 \approx c_5}_{\mathcal{K}_2} \succ \underbrace{c_6}_{\mathcal{K}_3}$$

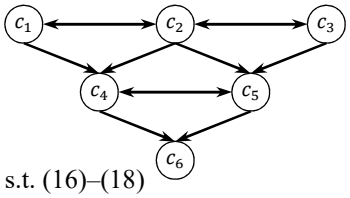


Figure 3 An optimal solution to the illustrative example min (15) s.t. (16)–(18)

An optimal solution to the problem:

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The linear ordering represented by the optimal solution P:

$$\underbrace{c_1}_{\mathcal{K}_1} \succ \underbrace{c_2}_{\mathcal{K}_2} \succ \underbrace{c_3}_{\mathcal{K}_3} \succ \underbrace{c_4}_{\mathcal{K}_4} \succ \underbrace{c_5}_{\mathcal{K}_5} \succ \underbrace{c_6}_{\mathcal{K}_6}$$

Figure 4 An integer optimal solution to the simple illustrative example min (19) s.t. (20)–(22)

for all $i, j \in \mathcal{K}_r$ for $r = 1, 2, \dots, R$, and let $i \succ j$ for all $i \in \mathcal{K}_r$ and for all $j \in \mathcal{K}_s$ for $r = 1, 2, \dots, R-1$ and for $s = r+1, r+2, \dots, R$.

5 An illustrative example

Let the alo-group $\mathcal{G} = (G, \odot, \leq)$ be the usual multiplicative group $\mathcal{R}_+ = (\mathbb{R}_+, \cdot, \leq)$ of the field of the reals with the usual multiplication and usual linear ordering, and with the neutral element $e = 1$. We are given $n = 6$ objects $c_1, c_2, c_3, c_4, c_5, c_6$ to be judged with respect to some criterion by $m = 2$ independent decision makers (evaluators or experts). The experts have independently assessed the relative importance of the two items in each pair of the objects with respect to the criterion. Matrices A_1 and A_2 presented in Figures 1 and 2, respectively, present the opinions of the two experts. Both matrices are coherent. The corresponding coherent priority vectors of the matrices are, e.g. $w^1 = \left(\frac{3}{10}, \frac{2}{10}, \frac{2}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}\right)$ and $w^2 = \left(\frac{3}{14}, \frac{3}{14}, \frac{3}{14}, \frac{2}{14}, \frac{2}{14}, \frac{1}{14}\right)$, respectively. We can see the opinions of the experts are different: while the 1st expert considers elements c_2 and c_3 to be less important that element c_1, the 2nd expert considers all three elements c_1, c_2, c_3 to be equally important; while the 1st expert considers elements c_4, c_5, c_6 to be equally important, the 2nd expert considers elements c_4 and c_5 to be more important that element c_6. Now, our purpose is to find a single consensual priority vector $v = (v_1, v_2, v_3, v_4, v_5, v_6) \in \mathbb{R}_+^6$. To this end, we find an integer optimal solution to the above given linear programming problem to minimize (9) subject to either (5)–(7) or (10), (11), and (13). First, we construct the objective function; that is, the matrix of coefficients:

$$C = \begin{pmatrix} 5 & 4 & 4 & 3 & 3 & 3 \\ 4 & 3 & 3 & 2 & 2 & 2 \\ 3 & 2 & 2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 & 2 \end{pmatrix} + \begin{pmatrix} 5 & 5 & 5 & 2 & 2 & 0 \\ 4 & 4 & 4 & 1 & 1 & 1 \\ 3 & 3 & 3 & 0 & 0 & 2 \\ 2 & 2 & 2 & 1 & 1 & 3 \\ 1 & 1 & 1 & 2 & 2 & 4 \\ 0 & 0 & 0 & 3 & 3 & 5 \end{pmatrix} = \begin{pmatrix} 10 & 9 & 9 & 5 & 5 & 3 \\ 8 & 7 & 7 & 3 & 3 & 3 \\ 6 & 5 & 5 & 1 & 1 & 3 \\ 4 & 3 & 3 & 1 & 1 & 3 \\ 2 & 1 & 1 & 3 & 3 & 5 \\ 0 & 1 & 1 & 5 & 5 & 7 \end{pmatrix} \tag{14}$$

Next, we solve the non-smooth integer optimization problem (minimize (9) subject to (5)–(7)); that is,

$$\text{minimize} \quad \sum_{i=1}^{6} \sum_{j=1}^{6} c_{ij} p_{ij} \tag{15}$$

subject to

$$\sum_{i=1}^{6} p_{ij} = 1 \qquad \text{for} \quad j = 1, 2, 3, 4, 5, 6, \tag{16}$$

$$\sum_{j=1}^{6} p_{ij} \leq i \times \max\left\{0,\ 1 + k - \sum_{j=1}^{6} p_{i+k,j}\right\} \qquad \text{for} \quad k = 1, \dots, 6 - i \qquad \text{for} \quad i = 1, 2, 3, 4, 5, 6, \tag{17}$$

$$p_{ij} \in \{0, 1\} \qquad \text{for} \quad i, j = 1, 2, 3, 4, 5, 6. \tag{18}$$

By using the Excel Solver, we find an optimal solution along with the respective consensual quasi-linear ordering of the set $\mathcal{C} = \{c_1, c_2, c_3, c_4, c_5, c_6\}$ presented in Figure 3.

We notice that the optimal solution $P = P_2$; that is, the opinion of the 2nd expert is consensual in this example. A consensual coherent priority vector is then, e.g., $v = w^2 = \left(\frac{3}{14}, \frac{3}{14}, \frac{3}{14}, \frac{2}{14}, \frac{2}{14}, \frac{1}{14}\right)$.

We also solve the simple linear programming problem (minimize (9) subject to (10), (11), and (13)); that is,

$$\text{minimize} \quad \sum_{i=1}^{6}\sum_{j=1}^{6} c_{ij}p_{ij} \tag{19}$$

subject to

$$\sum_{i=1}^{6} p_{ij} = 1 \qquad \text{for} \quad j = 1,2,3,4,5,6, \tag{20}$$

$$\sum_{j=1}^{6} p_{ij} = 1 \qquad \text{for} \quad i = 1,2,3,4,5,6, \tag{21}$$

$$p_{ij} \geq 0 \qquad \text{for} \quad i,j = 1,2,3,4,5,6. \tag{22}$$

By using the Excel Solver, we find an optimal solution along with the respective consensual linear ordering of the set $\mathcal{C} = \{c_1, c_2, c_3, c_4, c_5, c_6\}$ presented in Figure 4.

A consensual coherent priority vector is then, e.g., $v = \left(\frac{6}{21}, \frac{5}{21}, \frac{4}{21}, \frac{3}{21}, \frac{2}{21}, \frac{1}{21}\right)$.

6 Concluding remarks

In Section 4, we proposed the notion of consensuality of two priority vectors and also that of two quasi-linear orderings of n objects $c_1, c_2, \dots, c_n$. Subsequently, given two quasi-linear orderings, we proposed the notion of grade of their non-consensuality. Finally, given m quasi-linear orderings $\succcurlyeq^1, \succcurlyeq^2, \dots, \succcurlyeq^m$, i.e. evaluations of the n objects by m evaluators (experts) with respect to some criterion, and yet an eventually consensual quasi-linear ordering $\succcurlyeq$ of the objects, we defined the total non-consensuality of $\succcurlyeq$ with $\succcurlyeq^1, \succcurlyeq^2, \dots, \succcurlyeq^m$ by formula (9), which is to be minimized. When constructing the objective function (9), we assumed that each of the m evaluators is equally skilled to judge the n objects $c_1, c_2, \dots, c_n$. Instead, we can express the importance and/or qualifications of the evaluators by weights $u_1, u_2, \dots, u_m > 0$ such that $u_1 + u_2 + \cdots + u_m = 1$. Then, using the general weighted power mean with $q \in (-\infty, +\infty) \setminus \{0\}$, of which the weighted arithmetic mean is a special case ($q = 1$), or the weighted geometric mean ($q = 0$), we can define the total non-consensuality of $\succcurlyeq$ with $\succcurlyeq^1, \succcurlyeq^2, \dots, \succcurlyeq^m$ as:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{k=1}^{m} u_k \times \left(\sum_{i^k=1}^{n} |i^k - i| \times p_{i^k j}^{k}\right)^q\right]^{1/q} \times p_{ij} \qquad \text{or} \qquad \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\prod_{k=1}^{m}\left(\sum_{i^k=1}^{n} |i^k - i| \times p_{i^k j}^{k}\right)^{u_k}\right] \times p_{ij}\,.$$

Acknowledgements

This work was supported by the Czech Science Foundation under grant number GAČR 21-03085S.

References

[1] Bana e Costa, C. A., & Vansnick, J.-C. (2008). A critical analysis of the eigenvalue method used to derive priorities in AHP. *European Journal of Operational Research, 187*(3), 1422–1428.

[2] Bartl, D., & Ramík, J. (2022). A new algorithm for computing priority vector of pairwise comparisons matrix with fuzzy elements. *Information Sciences, 615*, 103–117.

[3] Bartl, D., & Ramík, J. (2023). An Algorithm to Compute a Joint Priority Vector of Pairwise Comparison Matrices with Fuzzy Elements in Group Decision Making. In *International Conference on Decision Making for Small and Medium-Sized Enterprises* (pp. 18–25). Karviná: Silesian University in Opava, School of Business Administration in Karviná.

[4] Cavallo, B., & D'apuzzo, L. (2009). A General Unified Framework for Pairwise Comparison Matrices in Multicriterial Methods. *International Journal of Intelligent Systems, 24*(4), 377–398.

[5] D'Apuzzo, L., Marcarelli, G., & Squillante, M. (2007). Generalized Consistency and Intensity Vectors for Comparison Matrices. *International Journal of Intelligent Systems, 22*(12), 1287–1300.

[6] Kułakowski, K. (2015). Notes on order preservation and consistency in AHP. *European Journal of Operational Research, 245*(1), 333–337.

[7] Ramík, J. (2015). Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences, 297*, 236–253.

[8] Ramík, J. (2020). *Pairwise Comparisons Method: Theory and Applications in Decision Making.* Cham: Springer.

[9] Saaty, T. L. (1980). *Analytic Hierarchy Process.* New York: McGraw-Hill.

[10] Saaty, T. L., & Vargas, L. G. (1984). Comparison of Eigenvalue, Logarithmic Least Squares and Least Squares Methods in Estimating Ratios. *Mathematical Modelling, 5*(5), 309–324.

Three-Parameter Lognormal Curves Combined with the Quantile Method of Parameter Estimation as Models of Salary Distribution

Diana Bílková[1]

Abstract. The main objective of this paper is the construction of salary distribution models through three-parameter lognormal curves in combination with the quantile method of point parameter estimation. Salary distribution models are constructed separately for men and women and according to educational attainment. An important aim is to capture the development of salary distributions over time during the period 2014–2020 and the specification of the typical shape of these models for individual categories of employees. The data for this research was taken from the Czech Statistical Office.

Keywords: three-parameter lognormal curve; quantile parameter estimation method; salary distribution model

JEL Classification: C51, C55, E24, J31, I31

AMS Classification: 60E05, 62F10, 62P20, 91B39

1 Introduction

The main purpose of this paper is to present the construction of models for the distribution of salaries of men and women in the Czech Republic, separated according to the educational attainment of the employee. Salary distribution models were constructed for the period 2014–2020, and one of the aims is to capture the development of salary distribution models over time for individual differentiated categories of employees. Salary distribution models were constructed using three-parametric lognormal curves, while the beginning of these curves was the amount of the minimum wage in the Czech Republic on January 1 of the previous year, as institutions do not respond to increases in the minimum wage quite flexibly, as is evident from the data. The remaining two parameters were estimated using the quantile method of point parameter estimation. An important task is to specify the typical shapes of salary distribution models for individual differentiated categories of employees. For example, publications [1], [2], [3] deal with the issue of the lognormal distribution and the associated estimation of its parameters. The data for this research come from the Czech Statistical Office.

2 Methodology

2.1 Three-Parameter Lognormal Distribution and Quantile Method

The random variable X has a three-parameter lognormal distribution with parameters μ, σ^2 and θ, where $-\infty < \mu < \infty$, $\sigma^2 > 0$, $-\infty < \theta < \infty$, if its probability density has the form$-\infty < \theta < \infty$,

$$\begin{aligned} f(x;\, \mu, \sigma^2, \theta) \quad &= \frac{1}{\sigma \cdot (x-\theta) \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{[\ln(x-\theta) - \mu]^2}{2\sigma^2} \right], \qquad x > \theta, \\ &= 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{else.} \end{aligned} \tag{1}$$

If the random variable X has a three-parameter lognormal distribution with parameters μ, σ^2 and θ, then the random variable

$$Y = \ln(X - \theta) \tag{2}$$

has a normal distribution with parameters μ and σ^2 and is a random variable

$$U = \frac{\ln(X - \theta) - \mu}{\sigma} \tag{3}$$

[1] Prague University of Economics and Business / Faculty of Informatics and Statistics, Department of Statistics and Probability, Sq. W. Churchill 1938/4, 130 67 Prague 3, bilkova@vse.cz.

has a standardized normal distribution. The parameter μ represents the expected value of the random variable (2) and the parameter σ^2 is the variance of the random variable (2). The parameter θ represents the beginning (theoretical minimum) of the three-parameter lognormal distribution of the random variable *X*. As mentioned above, the random variable (2) has a normal distribution with parameters μ and σ^2. The 100**P*% quantile of the normal distribution with parameters μ and σ^2 of the random variable *Y* has the form

$$y_P = \mu + \sigma u_P, \tag{4}$$

where u_P is the 100**P*% quantile of the standardized normal distribution.

We consider the value of the theoretical minimum θ to be the amount of the monthly minimum wage valid on January 1 of the previous year, as institutions do not react to the increase of the minimum wage quite flexibly, as is clear from the data. We therefore assume that we know the value of the parameter θ and estimate the value of the remaining two parameters μ and σ^2. For estimation, we use the sample 50% and 75% quantiles (the sample median and sample upper quartile), which we calculate from the sample data set

$$\tilde{x}_{50} \text{ a } \tilde{x}_{75},$$

which we substitute into the equations

$$y_{0,50} = \ln(\tilde{x}_{0,50} - \theta) = \hat{\mu} + \hat{\sigma} u_{0,50} = \hat{\mu}, \tag{5}$$

$$y_{0,75} = \ln(\tilde{x}_{0,75} - \theta) = \hat{\mu} + \hat{\sigma} u_{0,75}. \tag{6}$$

We substitute equation (5) into equation (6) and after adjustment we get

$$\frac{\ln(\tilde{x}_{0,75} - \theta) - \hat{\mu}}{u_{0,75}} = \frac{\ln(\tilde{x}_{0,75} - \theta) - \ln(\tilde{x}_{0,50} - \theta)}{u_{0,75}} = \hat{\sigma}. \tag{7}$$

We estimate the remaining parameters μ and σ^2 of the four-parameter lognormal distribution using formulas

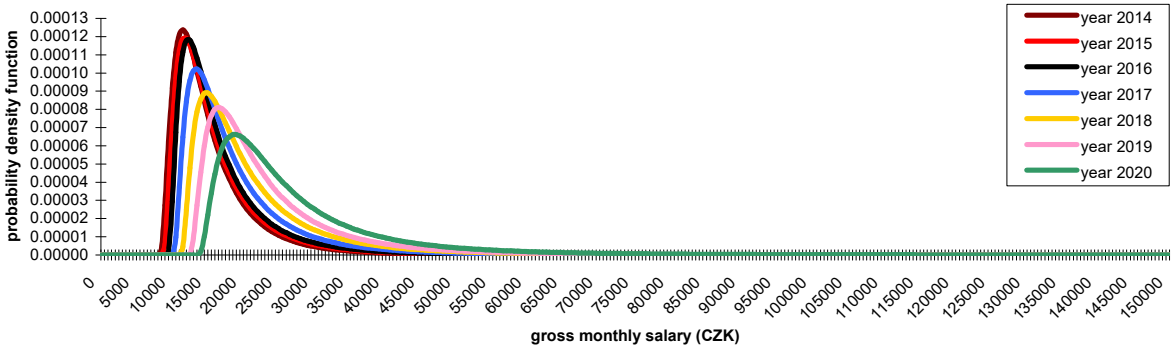
$$\hat{\mu} = \ln(\tilde{x}_{0,50} - \theta), \tag{8}$$

$$\hat{\sigma} = \frac{\ln(\tilde{x}_{0,75} - \theta) - \ln(\tilde{x}_{0,50} - \theta)}{u_{0,75}}. \tag{9}$$

3 Results

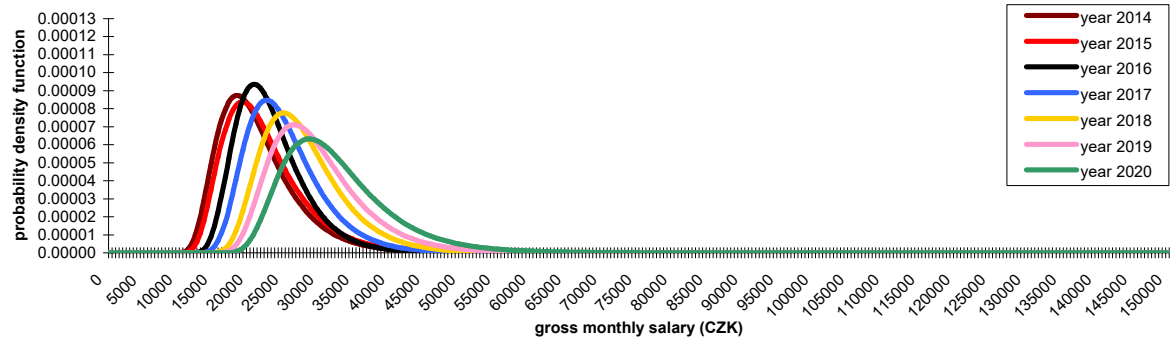
Source: Own calculation, own construction

Figure 1 Development of the model distribution of the gross monthly salary of men in the period 2014–2020 for the category primary and incomplete education



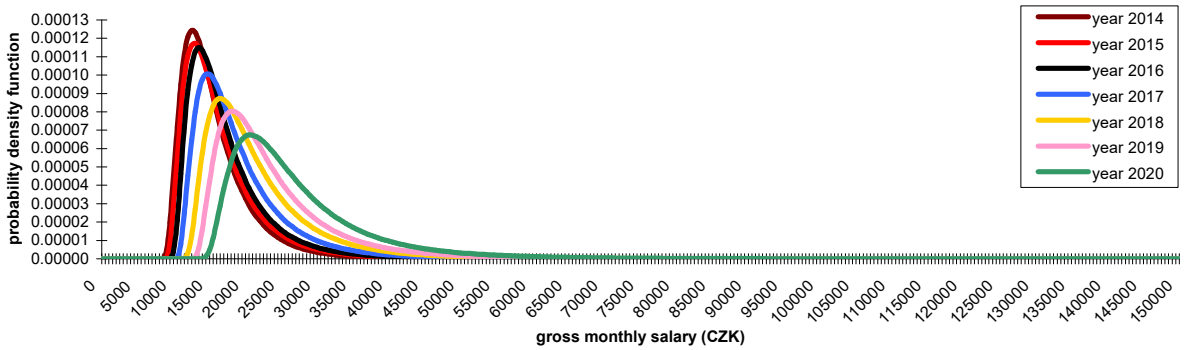
Source: Own calculation, own construction

Figure 2 Development of the model distribution of the gross monthly salary of women in the period 2014–2020 for the category primary and incomplete education



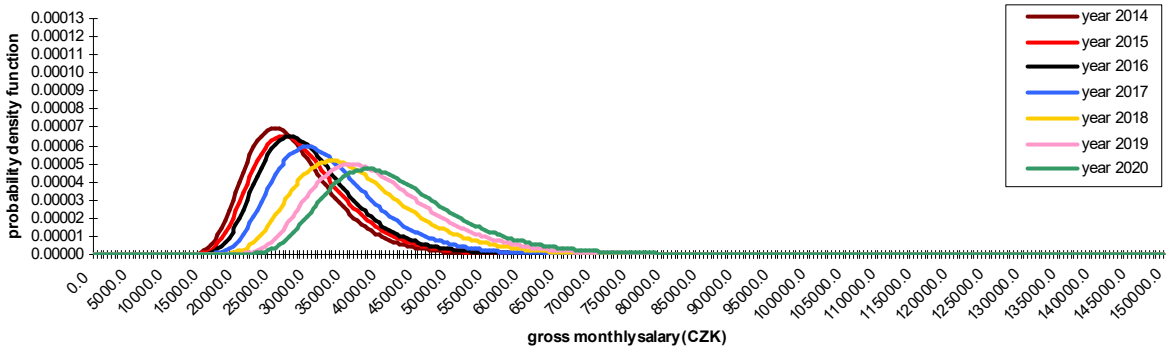
Source: Own calculation, own construction

Figure 3 Development of the model distribution of the gross monthly salary of men in the period 2014–2020 for the category secondary education without A-level examination



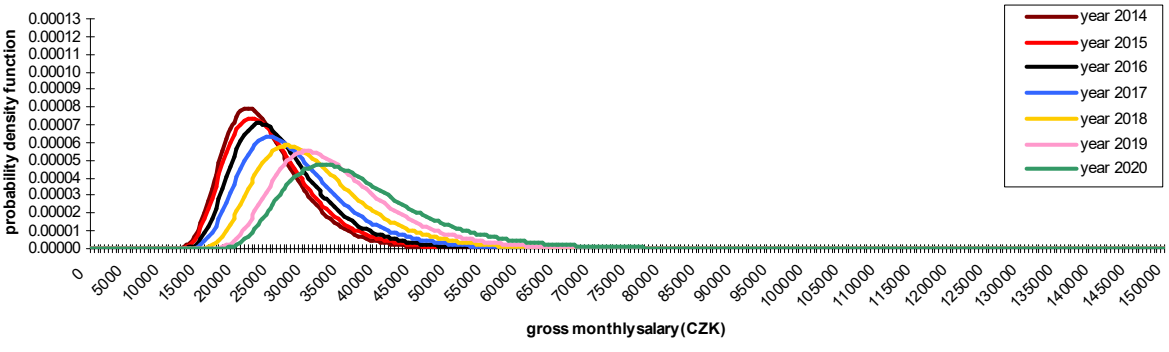
Source: Own calculation, own construction

Figure 4 Development of the model distribution of the gross monthly salary of women in the period 2014–2020 for the category secondary education without A-level examination



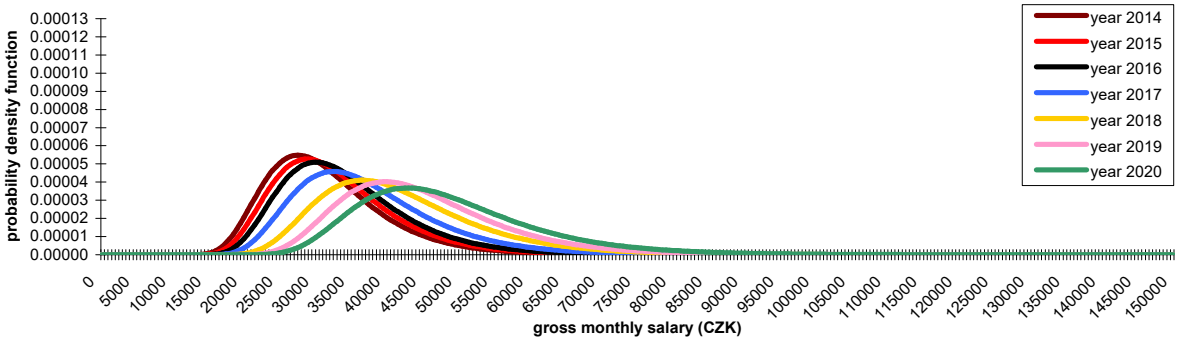
Source: Own calculation, own construction

Figure 5 Development of the model distribution of the gross monthly salary of men in the period 2014–2020 for the category secondary education with A-level examination



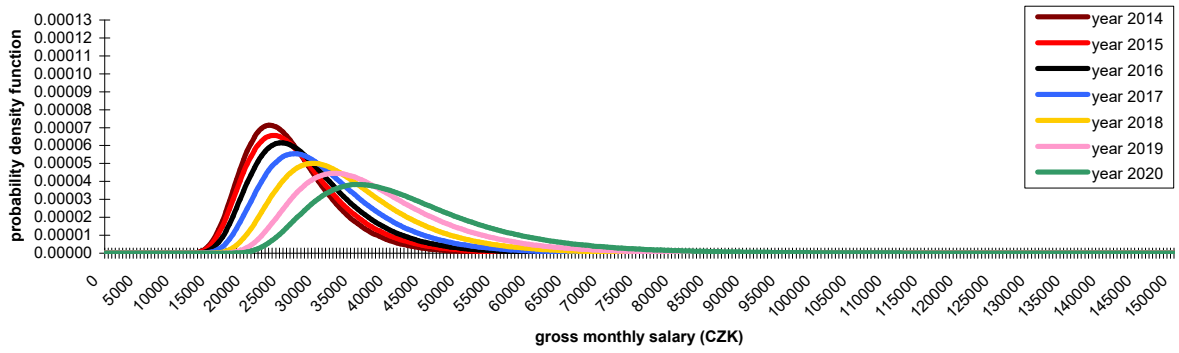
Source: Own calculation, own construction

Figure 6 Development of the model distribution of the gross monthly salary of women in the period 2014–2020 for the category secondary education with A-level examination



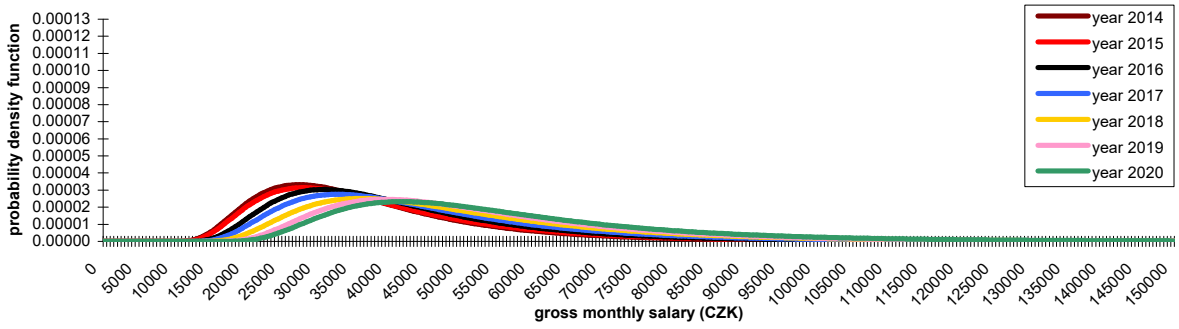
Source: Own calculation, own construction

Figure 7 Development of the model distribution of the gross monthly salary of men in the period 2014–2020 for the category post-secondary non-tertiary and bachelor´s education



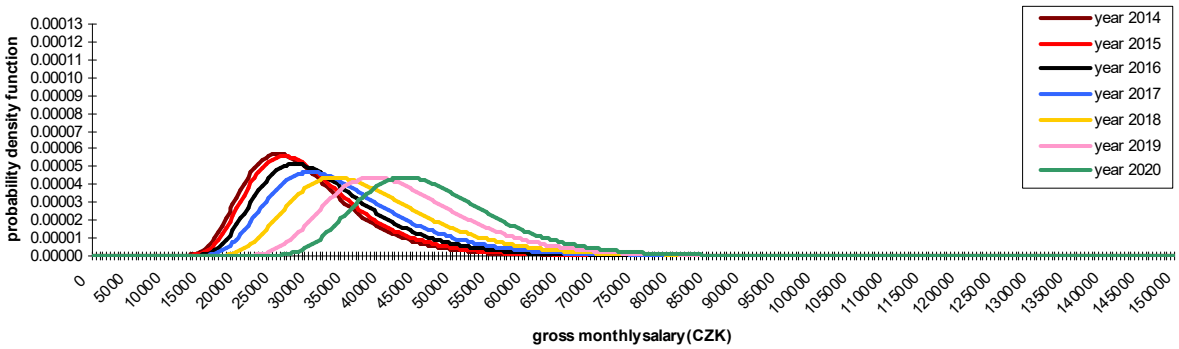
Source: Own calculation, own construction

Figure 8 Development of the model distribution of the gross monthly salary of women in the period 2014–2020 for the category post-secondary non-tertiary and bachelor´s education



Source: Own calculation, own construction

Figure 9 Development of the model distribution of the gross monthly salary of men in the period 2014–2020 for the category higher education



Source: Own calculation, own construction

Figure 10 Development of the model distribution of the gross monthly salary of women in the period 2014–2020 for the category higher education

		Parameter estimation				
			Men		Women	
Year	**Educational attainment**	**θ**	**μ**	**σ**	**μ**	**σ**
2014	Primary and incomplete	8,000	9.214 623	0.516 697	8.669 110	0.717 034
	Secondary without A-level examination	8,000	9.396 303	0.412 907	8.780 147	0.585 497
	Secondary with A-level examination	8,000	9.861 559	0.313 614	9.660 115	0.340 149
	Post-secondary non-tertiary and bachelor´s	8,000	10.005 020	0.349 750	9.746 536	0.347 388
	Higher	8,000	10.162 920	0.535 775	9.916 819	0.367 100
2015	Primary and incomplete	8,500	9.256 309	0.504 848	8.674 850	0.763 623
	Secondary without A-level examination	8,500	9.427 888	0.417 951	8.786 161	0.635 357
	Secondary with A-level examination	8,500	9.903 513	0.320 580	9.675 944	0.362 149
	Post-secondary non-tertiary and bachelor´s	8,500	10.042 270	0.350 428	9.767 267	0.373 298
	Higher	8,500	10.190 620	0.557 157	9.937 979	0.370 145
2016	Primary and incomplete	9,200	9.216 090	0.555 986	8.663 441	0.801 042
	Secondary without A-level examination	9,200	9.464 504	0.351 709	8.793 890	0.649 938
	Secondary with A-level examination	9,200	9.919 279	0.317 923	9.710 429	0.365 647
	Post-secondary non-tertiary and bachelor´s	9,200	10.071 010	0.351 995	9.802 339	0.385 923
	Higher	9,200	10.261 140	0.530 203	9.992 372	0.381 547
2017	Primary and incomplete	9,900	9.306 484	0.583 812	8.808 301	0.809 699
	Secondary without A-level examination	9,900	9.556 527	0.354 706	8.918 806	0.659 639
	Secondary with A-level examination	9,900	10.006 390	0.319 681	9.775 069	0.386 409
	Post-secondary non-tertiary and bachelor´s	9,900	10.156 920	0.360 247	9.881 374	0.397 633
	Higher	9,900	10.336 810	0.545 608	10.067 730	0.389 455
2018	Primary and incomplete	11,000	9.395 754	0.559 580	8.935 618	0.833 107
	Secondary without A-level examination	11,000	9.653 529	0.350 381	9.056 771	0.664 399
	Secondary with A-level examination	11,000	10.122 480	0.328 011	9.858 911	0.385 269
	Post-secondary non-tertiary and bachelor´s	11,000	10.275 190	0.357 307	9.973 278	0.402 980
	Higher	11,000	10.428 500	0.546 078	10.167 710	0.372 397
2019	Primary and incomplete	12,200	9.460 385	0.606 119	9.038 200	0.815 355
	Secondary without A-level examination	12,200	9.689 533	0.372 667	9.137 841	0.666 144
	Secondary with A-level examination	12,200	10.185 100	0.318 233	9.944 410	0.372 998
	Post-secondary non-tertiary and bachelor´s	12,200	10.340 270	0.339 981	10.074 310	0.410 030
	Higher	12,200	10.485 240	0.519 425	10.300 690	0.320 242
2020	Primary and incomplete	13,350	9.566 904	0.628 739	9.245 378	0.802 150
	Secondary without A-level examination	13,350	9.769 275	0.388 726	9.325 984	0.651 365
	Secondary with A-level examination	13,350	10.230 880	0.321 949	10.023 390	0.404 496
	Post-secondary non-tertiary and bachelor´s	13,350	10.412 230	0.346 647	10.192 550	0.427 873
	Higher	13,350	10.531 670	0.526 432	10.396 240	0.288 968

Source: Own calculation

Table 1 Parameter estimations of the four-parameter lognormal distribution model

Table 1 contains the estimation of the parameters of the three-parametric lognormal curves obtained by the quantile method. The development of salary distribution models during the monitored period is shown in the figures 1–10.

From the above figures, it is clear that for each distinguished category of employees, salary distribution models always show a higher skewness and kurtosis at the beginning of the monitored period, with a lower level and variability at the same time. Over time, salary distribution models become less skewed, they have less kurtosis, and their level and variability increase.

The salary distribution of women has higher skewness and kurtosis with lower level and variability than the salary distribution of men in the same year and corresponding educational attainment category.

The lower the level of educational attainment of employees, the higher the skewness and kurtosis of the salary distribution, and at the same time the lower the level and variability. As the level of educational attainment increases, the skewness and kurtosis of salary distributions decrease, and their level and variability increase.

However, the question of the appropriateness of a given curve for a salary distribution model is not quite a common mathematical-statistical problem in which we test the null hypothesis

$$H_0\text{: The sample comes from an assumed theoretical distribution}$$

against the alternative hypothesis

$$H_1\text{: non } H_0,$$

for in goodness-of-fit tests in cases of salary distribution models we often work with large samples, and therefore the test would almost always lead to the rejection of the null hypothesis. This results not only from the fact that, with such large sample ranges, the power of the test is so great at the chosen significance level that the test will reveal even the smallest deviations of the actual salary distribution and model, but also from the very principle of test construction. However, we are not interested in such small deviations, only an approximate agreement of the model with reality suffices.

4 Conclusion

Models of the salary distribution of men and women according to the category of educational attainment were constructed using three-parameter lognormal curves and the quantile method of point estimation of parameters. The development of salary distribution models over time was captured and the typical shapes of salary distributions for individual separated categories of employees were identified.

For future research, it is possible to go in the direction of constructing models of salary distribution of men and women and the same educational attainment categories on the same data through four-parameter lognormal curves, again using the quantile method of point estimation of parameters. Comparing the accuracy and shapes of three-parameter and four-parameter lognormal models of the salary distribution can yield interesting findings.

Acknowledgements

This paper was subsidized by the funds of institutional support of a long-term conceptual advancement of science and research number IP400040 at the Faculty of Informatics and Statistics, University of Economics and Business, Czech Republic.

References

[1] Cohen, A. C. & Whitten, B. J. (1980). Estimation in the Three-Parameter Lognormal Distribution. *Journal of American Statistical Association*, *75(370)*, 399–404.

[2] Giesbrecht, F. & Kempthorne, O. (1976). Maximum Likelihood Estimation in the Three-Parameter Lognormal Distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38(3)*, 257–264.

[3] Singh, V. P. (1998). Three-Parameter Lognormal Distribution. *Entropy-Based Parameter Estimation in Hydrology*, *30*, 82–107.

Strategic Management DSS Preferences Evaluation Method Using ANP – Application of Behavioral Economics

Petra Bláhová[1], Jan Rydval[2], Helena Brožová[3]

Abstract. The increasing volume of information and rapid digitalization creates a constantly changing environment where the ability to make effective and timely decisions becomes crucial. Enabled by the development of information technologies, decision support systems (DSS) became an essential tool in this environment. Although the DSS is known to improve organizational performance and success, strategic management DSS use is less than optimal, and intuitive expert estimation and knowledge is preferred. To improve the level of DSS use and DSS acceptance, the paper proposes a method for evaluating senior managers' preferences for DSS capabilities in the context of behavioral economics (BE) heuristics and cognitive bias, and in the context of selected strategies. Proposed method is based on the Analytic Network Process (ANP) suitable for complex decision-making problems, with multiple criteria and alternatives considering the influence of objective and subjective factors.

This paper uses the ANP model to evaluate DSS benefits and capabilities in terms of both decision elements and cognitive bias. The proposed model is flexible regarding selected elements, allowing application in any organization that needs to evaluate its management decision style and the complexity of managers' preferences before proposing the DSS framework. Sensitivity analysis results seem to be an essential part of increasing the understanding of the success of DSS acceptance as such analysis allows to find out how the importance of each element of the DSS changes as the preferences of each element of the ANP decision model increase.

Keywords: ANP, DSS preferences evaluation, strategic management decisions, sensitivity analysis, BE, cognitive bias

JEL Classification: C44, D91
AMS Classification: 90B50

1 Introduction

Access to constantly changing information is creating an urgency to make more frequent decisions. The decision frequency is increasing on all management levels in all organizations. While strategic decisions have a significant impact on the success of an organization, the effective and timely execution of such decisions is crucial. Although strategic decisions have a significant impact, and the effectiveness of strategic decisions greatly depends on decision-making methods based on integrated automation and informatization [1], DSS use in strategic management is limited [2]. Integrated automation and informatization, which refers to integrating automated systems and using information technology, can play a vital role in supporting the decision-making process. But according to Thaler [23], the importance of decision grows, and the tendency to rely on quantitative analysis declines. In the current environment this represents a severe problem for organizations as, according to Simon's bounded rationality theory [22], the decision makers could not have perfect knowledge of a decision situation and, further, they are limited in their cognitive and information processing abilities.

The importance and benefits of DSS are generally understood; however, the impact of DSS in strategic decisions in the context of BE dual theory and cognitive bias is limited. Instead, much of the research focuses on specific DSS capabilities' impacts. For instance, such as task motivation to use DSS [7], business simulation games where

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, blahovap@pef.czu.cz
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, rydval@pef.czu.cz
[3] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, brozova@pef.czu.cz

correct and defective DSS are analysed in terms of trust in automation and intention to use [6]. Frequent capabilities in recent research: *Explanation* capability is covered by several authors, such as how different explanation treatments cause a revision of the initial decision [13], and how explanation length affects confidence level [10]. *Cooperation* capability is covered, for example: how shared knowledge creation effect strategies [8] [26]. *Experience* capability *is* covered by how past experience improves business processes by [9], and relevant prior experience information increases confidence by [24]. Often DSS capabilities are evaluated in context to their impacts on decision-making elements, such as the *Certainty* of a decision-maker where prior experience with a given decision problem resulted in the highest certainty [12]. *The quality* of a decision is included as a fundamental DSS-related impact, according to Power [16]. *The efficiency* of a decision-making process research examines the automated DSS effect on task performance or DSS impact on organizational intelligence [12], [14]. A comprehensive overview of recent DSS research with BE context is provided by Arnott [3] [4]. Our research combines the current results in evaluating DSS capabilities and impacts in the context of the behavioral economy by adding the elements of cognitive biases into the proposed evaluation method.

The paper's main aim is to evaluate senior managers' preferences for DSS capabilities in the context of impacts on decision-making elements, behavioral economics (BE) cognitive bias and selected business strategies in a pilot study. Evaluation of DSS capabilities and related effects focusing on the corporate environment represents a complex relation problem. Not all quantitative methods are suitable for a survey on a corporate level. There needs to be a level of automation and the ability to cover a complex problem. Regarding complex relations among DSS capabilities, impacts, cognitive biases, and business strategies, the ANP method is used.

2 Materials and Methods

2.1 Decision Support Systems

DSS implementation reflects an organization's specific needs and often represents a significant investment. Specifically, tactical and strategic managers are the DSS users, and therefore evaluation of their preferences the DSS capabilities is crucial for successful DSS implementation and use. Five DSS capabilities were chosen for our pilot study based on the most recent literature research, which included capabilities' impacts on decision elements. Selected DSS capability elements represent the decision alternatives cluster in the model, and their definitions are provided in Table 1, while referenced in this paper's introduction. Discovery and analysis of the complexity in cognitive evaluation are necessary for successful DSS use. The ANP model proposed in our study reflects such complexity, and several criteria clusters are introduced in ANP model construction chapter.

Cluster	Elements	Definition
DSS Capabilities (alternatives cluster)	Explanations	System provides supplemental explanations alongside recommended alternatives
	Cooperation	System enables cooperation with others, sharing knowledge, transparent visibility
	Past Experience	System recommendations are using prior similar decisions, imitating experience
	Openness	System information is transparent and open to all users without restrictions
	Scenario Simulation	System includes scenario simulations module based on actual data

Table 1 DSS capabilities – decision alternatives

2.2 Analytic Network Process

The Analytic Network Process (ANP) is a decision-making process that allows complex decisions across a wide range of decision-making problems. The ANP is based on pairwise comparisons, which is a process of comparing pairs of items to judge which elements of each pair are preferred or have a more significant amount of some quantitative property. The ANP considers multiple decision factors and manages the decision-making problem's structure as a network. The ANP is a generalization of the Analytic Hierarchy Process (AHP) because it includes the relationships between the elements of the problem structure hierarchy [20], i.e., the problem structures has interdependencies between the elements of the same level of the hierarchy and whole problem structure is composed like a network. Many decision problems cannot be structured hierarchically, i.e., they cannot be structured into an AHP model (see [19] for introduction to AHP theory), because they involve many interactions and dependencies of higher-level elements in a hierarchy on lower-level elements and between the same-level elements. Therefore, ANP is represented by a network rather than a hierarchy and is more suitable for complex decision-making problems ([20], [19], [21]).

The basic elements of the ANP method are as follows:

1) Creation of a decision network of the decision-making problem structure. The decision network describes the outer dependence among different sets of elements arranged in clusters.

2) Pairwise comparisons matrix. Pairwise comparisons of the decision elements within and among the clusters are performed according to their relevance to the decision-maker [20]. The consistency of these comparisons has to be controlled. The rate of consistency is measured by a consistency ratio *CR* defined by Saaty:

$$CR = \frac{CI}{RI_n} \tag{1}$$

where RI_n represents the random inconsistency index. The *CI* stands for consistency index, and for a comparisons matrix, is calculated as:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{2}$$

where λ_{max} is the largest eigenvalue of Saaty's comparisons matrix, and *n* is the number of criteria. A pairwise comparison matrix is considered inconsistent for CI>0.1, and CI>0.15 for large clusters with a number of elements > 10, respectively, according to Saaty [17]. In situations of inconsistency, the proposed method of Saaty's matrix modification optimization model of minimum deviation by Hlavaty [11] should be applied [5]. Supermatrix construction ([20], [19]). The priorities derived from the pairwise comparisons are entered into the appropriate position in the Unweighted Supermatrix. This Supermatrix has to be normalized using clusters weights, i.e., the Weighted Supermatrix is calculated.

3) Limit Matrix computation and global preferences of decision elements are obtained. The Limit Matrix is used to obtain stable preferences from Weighted Supermatrix. Raising the Weighted Supermatrix to powers generates the Limit Matrix. These preferences serve as the analysis of preferences of decision-making elements. See [20] for an introduction to the standard steps of the Limit Matrix calculation.

Sensitivity analysis in the ANP and SuperDecisions software

To ensure the stability of global preferences or the ranking of nods in the cluster of alternatives, the ANP model enables to conduct of a sensitivity analysis. The sensitivity analysis involves the elements with the highest weights and assesses their impact on all other decision elements or alternatives. The resulting data from the sensitivity analysis are interpreted by changing the input value of the selected node in the Unweighted Supermatrix from value 0 to 1 and computing the corresponding priorities of the alternatives from the Limit Supermatrix. ANP is in this paper carried out by the SuperDecisions software [18].

3 Results

A decision model is presented to support the decision of which system capability is preferred by senior management. The decision model includes a relational network of decision elements grouped in defined clusters.

3.1 ANP Model Construction

The decision model includes a relational network of individual decision elements grouped in defined clusters. Our model provides decision support for senior management from the perspective of individual decision elements based on pairwise comparisons. The decision-maker will determine each cluster's and node's preferences, and the model output advises the most important DSS capabilities. We have built a model that has five clusters in total. One main goal cluster, one cluster of alternatives (DSS Capabilities), and three clusters of criteria (Decision-making Elements, Cognitive Bias, and Business Strategies) were proposed. In this ANP model, the clusters are set to have equal weight. The ANP model network structure and the relationships are shown in Figure 1.

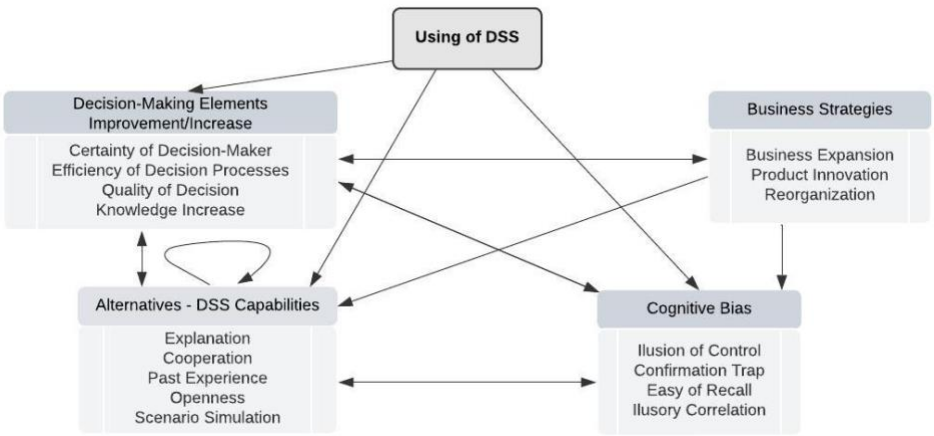


Figure 1 ANP model network structure

Clusters of criteria elements are described in Table 2. Decision-making Elements include essential decision-making elements where the significance of improvement is assessed in the model. Cluster Cognitive Bias includes cognitive biases selected based on Arnott's comprehensive list [3] and reduction/elimination significance is evaluated in terms of the use of DSS and the other criteria shown in our model. Here only four cognitive biases are selected, each representing the major bias category as defined by Arnott [3]. The last cluster includes business strategies, as the company business strategy reflects the external market environment leading to strategic decisions which impact DSS use. Business Strategies and Cognitive Bias elements shall be adapted for each specific company situation.

Clusters	Elements	Definition
Business Strategies	Business Expansion	The main strategy is business expansion - new market, products, or clients
	Product Innovation	Main strategy is product innovation - new development, service, or product
	Reorganization	Main strategy is reorganization and cost reduction - business turn-around
Cognitive Bias Reduction	Confirmation Trap	Tend to seek confirmation evidence and not for discomforting information
	Easy of Recall	Events can be overweighted if easily recalled, recency memory effect
	Illusion of Control	Poor decision may lead to good outcome - false feeling of control induced
	Illusory Correlation	Probability of 2 events occurring together overestimated if co-occured in the past
Decision-Making Elements Increase	Decision Certainty	Using DSS increases decision-maker certainty with his/her decision
	Processes Efficiency	System improves the decision-making process in terms of time and cost
	Knowledge Increase	Implemented and used DSS enables and promotes knowledge increase
	Quality of Decision	Using the system increases decision quality - adequacy, and accuracy

Table 2 Criteria clusters with elements definition

3.2 Pilot Study - ANP Model for DSS Capabilities Evaluation

The ANP model decision network is created to evaluate senior managers' preferences for DSS capabilities and benefits (decision-making elements improvements) in the context of behavioral economics (BE) cognitive bias and selected business strategies. The proposed model for this pilot study was evaluated during a face-to-face meeting with two senior executive managers in the roles of CEO and EVP, with more than 20 years of experience from a large global corporation. The normalized values of preferences of DSS capabilities alternatives from the Limit matrix are shown in Table 3.

Cluster	Decision Node	Preference	Cluster	Decision Node	Preference
DSS Capabilities Alternatives	Cooperation	0.1948	*Cognitive Bias Reduction*	Confirmation trap (Confirmation)	0.2152
	Explanation	0.1615		Easy of recall (Availability)	0.1955
	Openness	0.1517		Illusion of Control (Overconfidence)	0.2898
	Past Experience	0.2486		Illusory Correlation (Representativeness)	0.2995
	Scenario Simulation	0.2434	*Decision-Making Elements Increase*	Certainty of Decision-Maker	0.2855
Business Strategies	Business Expansion	0.2509		Efficiency of Decision Processes	0.1931
	Product Inovation	0.4370		Knowledge Increase	0.3330
	Reorganization	0.3121		Quality of Decision	0.1884

Table 3 ANP model limit matrix – final normalized preferences

In this pilot study, Past Experience is the most preferred with 24.86% weight, followed by the Scenario Simulation as the second most preferred DSS capability with 24.34% weight. In the case of DSS implementation, the organization shall focus on these capabilities. On the contrary, openness is the least important with 15.17% weight. By using this approach, DSS can be tailored to organization's environment. Importance of criteria elements provide valuable information about management decision culture. Most preferred elements are for example Product Innovation business strategy and increase of Knowledge is the most preferred/important positive DSS impact.

3.3 Sensitivity Analysis of Priorities of DSS Capabilities

Increasing understanding of the success of the DSS acceptance includes sensitivity analysis. The individual value of DSS capabilities (model alternatives) was tested by changing the preferences/importance of the model elements. By changing the preference/importance of the model element, the model alternatives' preferences change (direction and size). This responsiveness information is available in the model for all element changes to all alternatives. In this pilot study, only four elements were selected, representing all criteria clusters: one element from Decision-making Elements, one from Business Strategy, and two from Cognitive Bias Reduction to emphasize the importance of cognitive biases elimination. Results for selected elements are shown in Figure 2.

The first two charts in the upper part show the sensitivity analysis results for the change in the importance of Easy of Recall and Illusion of Control cognitive biases reduction. The increased importance of Easy of Recall bias reduction (reduction or even elimination using easily recalled memory events or readily available recent memory when making decisions) shows Past Experience preference sharp increase and a Scenario Simulation sharp decrease. The increased importance of the Illusion of Control bias reduction (eliminating false feeling of control when making wrong decisions with unrelated good results) shows the mild increase in Past Experience and Scenario Simulation. The lower part of Figure 2 shows the results for the Certainty of the Decision-maker where Cooperation importance sharply increases. Business Expansion change shows a low influence on all alternatives.

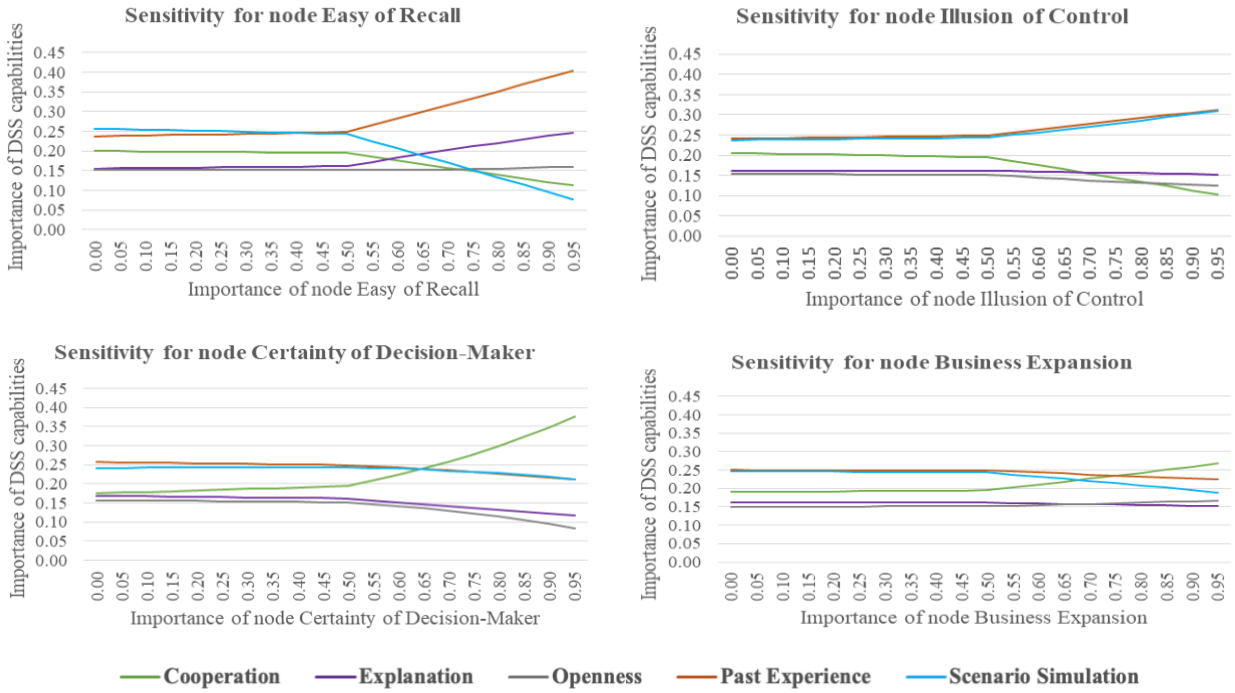


Figure 2 Sensitivity analysis for DSS capabilities

4 Discussion and Conclusion

ANP models usually use smaller-scale surveys or expert team information [15], [25]. Using ANP for evaluating larger-scale survey, the data processing may be complicated, especially in larger models. The presented ANP evaluation method assumes an ANP expert involvement to explain the ANP method and a strategic decision processes expert to guide management through the concepts of cognitive bias. The proposed model is not suitable for a short evaluation process. Instead, the method requires deep consideration to receive valuable results benefiting the organization uniquely. The weakness of this model is the source of the information obtained in a given organization. It is always necessary to include as many users of the future DSS as possible to avoid their subjectification. In our pilot study, Openness capability was evaluated as the least important, and this result demonstrates the potential less emphasis on information sharing and data transparency typical for senior management. Another weakness is achieving the required consistency. In our pilot study, achieving consistency 0.1 was difficult due to the model's complexity and the missing introduction to cognitive bias. Therefore, we need to find a better way of maintaining consistency, for example, the method introduced by Hlavaty and Brozova [11], and applied [5].

In conclusion, we proposed the model for evaluating DSS capabilities in this paper and obtained and assessed data from the pilot study. The model's suitability and reliability evaluation provided valuable findings for the subsequent research and future method applications. The evaluation also helped to identify weaknesses and improvements in the proposed model. Pilot study model results provide final DSS capabilities preferences and all model elements. The crucial part of the proposed model is sensitivity analysis, as it uncovers the relations between element preference change and capability preference change. Sensitivity analysis identified highly related alternatives and cognitive bias elements and indicated relevant results for the future research. Future research shall expand the pilot study by obtaining more data for potential method application case study.

References

[1] Alieinykov, I., Thamer, K.A., Zhuravskyi, Y., Sova, O., Smirnova, N., Zhyvotovskyi, R., et al. (2019). Development of a method of fuzzy evaluation of information and analytical support of strategic management. *Eastern-European Journal of Enterprise Technologies*. 6 (2–102), 16–27.

[2] Arnott, D. & Gao, S. (2019). Behavioral economics for decision support systems researchers. *Decision Support Systems*. 122 (May), 113063.

[3] Arnott, D. & Gao, S. (2022). Behavioral economics in information systems research: Critical analysis and research strategies. *Journal of Information Technology*, *37*(1), 80–117.

[4] Arnott, D. & Pervan, G. (2016). A critical analysis of decision support systems research. In: *Formul. Res. Methods Inf. Syst. Vol. 2*,.

[5] Blahova, P. & Brozova, H. (2022). DSS Capabilities Evaluation Using ANP with Methods for Consistency Improvement. in: *40th International Conference Mathematical Methods in Economics 2022 Proceedings 07.09.2022, Jihlava*. Jihlava: College of Polytechnics Jihlava, 2022. s. 21-26.

[6] Brauner, P., Philipsen, R., Calero Valdez, A. & Ziefle, M. (2019). What happens when decision support systems fail? — the importance of usability on performance in erroneous systems. *Behaviour and Information Technology*. 38 (12), 1225–1242.

[7] Chan, S.H., Song, Q., Sarker, S. & Plumlee, R.D. (2017). Decision support system (DSS) use and decision performance: DSS motivation and its antecedents. *Information and Management*. 54 (7), 934–947.

[8] Fan, S. & Shen, Q. (2011). The effect of using group decision support systems in value management studies: An experimental study in Hong Kong. *International Journal of Project Management*. 29 (1), 13–25.

[9] Ghattas, J., Soffer, P. & Peleg, M. (2014). Improving business process decision making based on past experience. *Decision Support Systems*. 59 (1), 93–107.

[10] Gönül, M.S., Önkal, D. & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*. 42 (3), 1481–1493.

[11] Hlavaty, R. & Brozova, H. (2022). Optimization Approach to Dealing with Saaty's Inconsistency. in: *40th International Conference Mathematical Methods in Economics 2022 Proceedings 07.09.2022, Jihlava*. Jihlava: College of Polytechnics Jihlava, 2022. s. 104-109.

[12] Langer, M., König, C.J. & Busch, V. (2021). Changing the means of managerial work: effects of automated decision support systems on personnel selection tasks. *Journal of Business and Psychology*. 36 (5), 751–769.

[13] Langer, M., König, C.J., Busch, V., Perçin, S., Gönül, M.S., Önkal, D., et al. (2020). Are you sure? Prediction revision in automated decision-making. *Decision Support Systems*. 68 (1), 1–19.

[14] Peixoto, L. de C., Golgher, A.B. & Cyrino, Á.B. (2017). Using information systems to strategic decision: an analysis of the values added under executive's perspective. *Brazilian Journal of Information Science*. 11 (2), 54–71.

[15] Perçin, S. (2008). Using the ANP approach in selecting and benchmarking ERP systems. *Benchmarking*. 15 (5), 630–649.

[16] Power, D.J. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Greenwood Publishing.

[17] Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*. 15 (3), 234–281.

[18] Saaty, T.L. (1996). *SuperDecisions Software for Decision-Making*. [Online]. Available at: https://www.superdecisions.com [cited 2023-04-20]

[19] Saaty, T.L. (1999). Fundamentals of the analytic network process. *Proceedings of the ISAHP 1999*. 1–14.

[20] *Saaty, T.L. (2001). Decision Making with Dependence and Feedback: The Analytic Network Process : the Organization and Prioritization of Complexity*. RWS Publications, Pittsburgh.

[21] Saaty, T.L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *European Journal of Operational Research*. 145 (1), 85–91.

[22] Simon, H. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*. 69 (1), 99–118.

[23] Thaler, R.H. (2015). *Misbehaving: The Making of Behavioral Economics*. W W Norton & Co.

[24] Waa, J. van der, Schoonderwoerd, T., Diggelen, J. van & Neerincx, M. (2020) Interpretable confidence measures for decision support systems. *International Journal of Human Computer Studies*. 144 (May), 102493.

[25] Xu, K. & Xu, J. (2020). A direct consistency test and improvement method for the analytic hierarchy process. *Fuzzy Optimization and Decision Making*. 19 (3), 359–388.

[26] Yang, C.W., Fang, S.C. & Lin, J.L. (2010). Organisational knowledge creation strategies: A conceptual framework. *International Journal of Information Management*. 30 (3), 231–238.

Portfolio Selection under Unstable Dynamics Using the Fuzzy Concept Operating with a Triplet of Moving Characteristics

Adam Borovička[1]

Abstract. Investment portfolio selection can be a very tortuous journey in which an investor faces a number of challenges. Asset prices can move very volatile. At the same time, it is usually possible to trace a certain trendiness in their behavior, which is due, among the other things, to their autocorrelation. The dynamics of development is not stable, it changes (unpredictably) over time. The phenomenon of such uncertainty is caused by the difficulty of predicting the evolution of prices which can be influenced by a variety of factors. To make a representative investment decision, the uncertainty, or an unstable dynamic environment must be taken into account. For this purpose, a methodological procedure integrating the fuzzy concept operating with a triplet of moving characteristics is proposed. Firstly, the continuity, trendiness is reflected by the moving elements, i.e. minimum, average and maximum of the returns. Secondly, the dynamics of the development is identified by means of an average indicator calculated as the triangular fuzzy number whose parameters are derived from the moving investment characteristics. Thirdly, the unstable dynamics, or uncertainty, is measured by the absolute deviation concept which practically quantifies the level of risk associated with the asset. The proposed concept of unstable dynamics is integrated into an investment portfolio making procedure relying on the fuzzified KSU-STEM method. The usefulness of the developed tool is demonstrated on the Czech open unit trust market. Portfolio can be formed according to the investor's preferences in a dynamically evolving environment, which significantly contributes to the adoption of a comprehensive, personally related investment decision.

Keywords: fuzzy, moving, portfolio selection, uncertainty, unit trust, unstable dynamics

JEL Classification: C44, C61, G11
AMS Classification: 90B50, 90C30

1 Introduction

Investing, investment decision making, is still a very fruitful and inspiring topic for many scientific papers. No wonder in a world of ever expanding (investment) opportunities. The world is becoming more crowded, more connected, more complex. This is no not different in the field of capital market. The range of investment instruments is ever expanding. The development of their prices is often very difficult to predict under the weight of many unstable, uncertain processes, sometimes with extreme manifestations.

Uncertainty, instability, is a capital market phenomenon that is caused by many factors. Uncertainty is usually integrated in models in the form of implicitly assumed probability distributions of asset returns with appropriate stable parameters, or characteristics. Examples include the well-known mean-variance model [10], or a number of more advanced (e.g. [3]) or multi-criteria approaches (e.g. [2]).

However, these models do not consider the dynamics of a price development, or its variability. The parameters, characteristics of the investment mostly change over time. The dynamics can be integrated by means of a fuzzy number representing the return (e.g. [6]). This is usually a triangular fuzzy number, which involves the basic question of determining its three parameters, for which average or extreme indicators are usually used. However, our research opens up another, often neglected, issue, namely the consideration of a certain trendiness in the behavior of prices or returns, which is accompanied, among other things, by their autocorrelation. Therefore, we propose to derive fuzzy number parameters from moving averages of returns that can capture this phenomenon. The comprehensive notion of performance is underlined by the inclusion of dynamic volatility, which is reflected

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

through an aggregate indicator of the variability of the particular parameters of the triangular fuzzy number, i.e. fuzzy return.

In order to make a representative investment decision under conditions of uncertainty, or better unstable dynamics, a proprietary multi-criteria methodological procedure is proposed. It uses a fuzzy concept operating with a triplet of moving characteristics (average, minimum and maximum) whose instability is expressed by the concept of absolute deviation [8]. The main application advantage of this concept is linearity, simplicity of calculation and illustrative interpretation. This approach is then integrated into the fuzzified multi-objective KSU-STEM method, which is able to take into account all relevant characteristics, circumstances, in portfolio making very efficiently [4]. Its robust implementation is demonstrated in the Czech market with increasingly popular open unit trusts. The composition of the resulting portfolio is analyzed, confronted with the output of a standard procedure that does not take into account volatile dynamics. Based on empirical research, the application-theoretical capabilities of the proposed approach are discussed.

The structure of the article is as follows. After the introduction, in Section 2, a methodological framework for making the investment portfolio under conditions of unstable dynamics is presented. The proposed procedure is applied to portfolio selection in Section 3. Finally, the results are discussed through an analysis with emphasis on the application-theoretical implications.

2 Methodological Framework for Unstable Dynamics Concept

Adequately capturing unstable (volatile) dynamics in investment decision-making requires the design of a comprehensive methodological framework. The emphasis will thus be on calculating the main characteristics of the investment - return and risk. Subsequently, the procedure of making an investment portfolio using the multi-objective optimization method will be described.

2.1 Investment Characteristics – Return and Risk

Capital appreciation is the essence of investing. Not surprisingly, then, **return** is an absolutely pivotal characteristic. Its calculation thus certainly deserves due attention. As mentioned above, the (expected) return is derived from its moving observation integrated into fuzzy expression in the form of a triangular fuzzy number. Let $r_{isp}, i = 1,2,...,m; s = 1,2,...,r;\ p = 1,2,...,q,$ denote the return of the *i*-th asset at the end of the *s*-th subperiod within the period *p*. The local average return in each period *p* can be easily calculate as

$$r_{ip}^{ave} = \frac{\sum_{s=1}^{r} r_{isp}}{r} \qquad i = 1,2,...,m;\ p = 1,2,...,q. \tag{1}$$

The expected return with each time period is supplemented by information on extreme values for a more comprehensive view of its dynamics. Let r_{ip}^{min}, and $r_{ip}^{max}, i = 1,2,...,m;\ p = 1,2,...,q,$ calculate the minimum, and maximum return of the *i*-th asset in terms of the *p*-th period.

$$r_{ip}^{min} = \min_{1 \le s \le r} r_{isp}, \qquad r_{ip}^{max} = \max_{1 \le s \le r} r_{isp} \qquad i = 1,2,...,m;\ p = 1,2,...,q\ . \tag{2}$$

Through overlapping periods (overlapping by one subperiod) we can observe the trendiness in behavior, or project the autocorrelation effect of the price or return of the assets. However, to make the investment decision, we rather need on representative indicator into which the partial views, volatile developments, are best integrated. Then the return of the *i*-th asset is proposed as the triangular fuzzy number $\tilde{r}_i = \left(r_i^{min}, r_i^{ave}, r_i^{max}\right)$, where the following holds

$$r_i^{min} = \frac{\sum_{p=1}^{q} r_{ip}^{min}}{q}, \qquad r_i^{ave} = \frac{\sum_{p=1}^{q} r_{ip}^{ave}}{q}, \qquad r_i^{max} = \frac{\sum_{p=1}^{q} r_{ip}^{max}}{q} \qquad i = 1,2,...,m\ . \tag{3}$$

The triplet of moving indicators thus helps to capture the dynamically evolving characteristics of an investment through a single aggregate indicator.

Return is depicted above as a dynamic element showing possible trend behavior. However, the movement of the price, or return, of an investment instrument may exhibit considerable volatility. The dynamics of the development over time are not invariable. This fact is reflected in the instability of the parameters of the triangular fuzzy number

which is quantified by the proposed risk measure. Then, the uncertainty of the parameters is measured by absolute deviation principle. Let us compute absolute deviation during the whole monitoring period for each parameter (3)

$$AD_i^{min} = \frac{\sum_{p=1}^{q} \left| r_{ip}^{min} - r_i^{min} \right|}{q}, \qquad AD_i^{ave} = \frac{\sum_{p=1}^{q} \left| r_{ip}^{ave} - r_i^{ave} \right|}{q}, \qquad AD_i^{max} = \frac{\sum_{p=1}^{q} \left| r_{ip}^{max} - r_i^{max} \right|}{q} \qquad i = 1, 2, ..., m \,. \tag{4}$$

A trio of indicators mapping the instability of the moving characteristics signaling their individual instability is the basis for a comprehensive quantification of the level of risk. The **risk**, and hence the manifestation of unstable dynamics, is quantified by aggregating the sub-indicators through the average as follows

$$AD_i = \frac{1}{3}\left(AD_i^{min} + AD_i^{ave} + AD_i^{max} \right) \qquad i = 1, 2, ..., m \,. \tag{5}$$

2.2 Portfolio Selection Procedure

In the next subsection, the investment decision making process is described with a dominant emphasis on the description of the methodology for investment portfolio selection.

At the beginning of the whole process, the investment policy needs to be set very precisely. This involves defining the purpose of the investment, the idea of the amount of expected return, the attitude to risk, the style of investment management, etc. On the basis of the specified investor profile, suitable investment instruments with appropriate characteristics are selected to potentially participate in the portfolio. The created portfolio is evaluated and, if necessary, revised over time.

To make a portfolio, the support tool based on the multi-objective programming method, can be used. This methodological concept, significantly inspired by (interactive) procedure proposed by Borovička [4, 5], can innovatively incorporate the phenomenon of unstable dynamics, or uncertainty in the capital market, as well as the individual preferences of the investor. The method is thus based on the user-friendly KSU-STEM approach [9], which is, however, fuzzified for a more comprehensive use, hence adapted to the investment situation. Its algorithm can be briefly described in the following few steps.

Step 1 The objectives and their weights are stated. The objective functions represent the observed characteristics of the investment. Their weights then express the investor's subjective preference for their importance. Let S_{of}^{min} and S_{of}^{max}, or F_{of}^{min} and F_{of}^{max} denote the set including the indices of the minimizing and maximizing objective functions with strict, or fuzzy coefficients. Further, let us introduce unifying sets $S_{of} = S_{of}^{min} \cup S_{of}^{max}$ and $F_{of} = F_{of}^{min} \cup F_{of}^{max}$, or $S_{of} \cup F_{of} = \{1, 2, ..., k\}$. Thus, *k* objective functions are formulated. Let $f_j(\mathbf{x}) = \mathbf{c_j^T x}$, $j \in S_{of}$, or $\tilde{f}_j(\mathbf{x}) = \tilde{\mathbf{c}}_\mathbf{j}^\mathbf{T}\mathbf{x}$, $j \in F_{of}$ denote the *j*-th objective function with strict $\mathbf{c_j^T} = (c_{1j}, c_{2j}, ..., c_{mj})$, or fuzzy coefficients $\tilde{\mathbf{c}}_\mathbf{j}^\mathbf{T} = (\tilde{c}_{1j}, \tilde{c}_{2j}, ..., \tilde{c}_{mj})$ with the vector of *m* variables $\mathbf{x} = (x_1, x_2, ..., x_m)^T$. The fuzzy coefficient of the *i*-th variable in the *j*-th objective function is formulated as a triangular fuzzy number $\tilde{\mathbf{c}}_{ij} = (c_{ij}^l, c_{ij}^m, c_{ij}^u), i = 1, 2, ..., m;\ j \in F_{of}$. Of course, the objective function is also a triangular fuzzy number $\tilde{f}_j(\mathbf{x}) = (f_j^l(\mathbf{x}), f_j^m(\mathbf{x}), f_j^u(\mathbf{x})) = (\mathbf{c_j^{l^T}x}, \mathbf{c_j^{m^T}x}, \mathbf{c_j^{u^T}x})$, where $\mathbf{c_j^{l^T}} = (c_{1j}^l, c_{2j}^l, ..., c_{mj}^l)$, $\mathbf{c_j^{m^T}} = (c_{1j}^m, c_{2j}^m, ..., c_{mj}^m)$ and $\mathbf{c_j^{u^T}} = (c_{1j}^u, c_{2j}^u, ..., c_{mj}^u)$. Finally, the weight of the *j*-th objective is denoted as w_j, $j = 1, 2, ..., k$. The weights are calculated using the user-friendly scoring method. They are therefore standardized as follows, $w_j > 0,\ j = 1, 2, ..., k, \sum_{j=1}^{k} w_j = 1$.

Step 2 The extreme, basal and ideal, values of the objectives are determined. The ideal value is formed as a minimum, or maximum of the minimizing, or maximizing objective function bound on all conditions forming the set of feasible solution X. Let $f_j(\mathbf{x_j^o}) = f_j^I$, $j \in S_{of}$, mark the ideal value of the *j*-th strict objective, where

$$\mathbf{x_j^o} = \arg\min_{\mathbf{x} \in X} f_j(\mathbf{x}) \quad j \in S_{of}^{min}, \qquad \mathbf{x_j^o} = \arg\max_{\mathbf{x} \in X} f_j(\mathbf{x}) \quad j \in S_{of}^{max}. \tag{6}$$

For an objective function with fuzzy coefficients, the ideal value is finding separately for each of the three functions determining the parameters of the triangular fuzzy number. Following (6), but solving three separate

problems, we obtain the ideal value of the *j*-th fuzzy objective function $\tilde{f}_j(\mathbf{x_j^o}) = \tilde{f}_j^I = (f_j^{lI}, f_j^{mI}, f_j^{uI}), j \in F_{of}$. The basal value of each objective function is determined in the context of the optimal solutions for other objectives. Then, let $f_j^B, j \in S_{of}$, or $\tilde{f}_j^B = (f_j^{lB}, f_j^{mB}, f_j^{uB}), j \in F_{of}$, denote the basal value of the *j*-th strict, or fuzzy objective function.

Step 3 The fuzzy goals are constructed which allows a problem with *k* objectives to be transformed to the one-objective model. Fuzzy goals thus mediate the simultaneous achievement of the best possible values of all objective functions under the investor preferences. The fuzzy goal is formulated as right-side, or left-side triangular fuzzy number for *j*-th minimizing, or maximizing strict objective as $\tilde{G}_j = (f_j^I, f_j^I, f_j^B), j \in S_{of}^{min}$, or $\tilde{G}_j = (f_j^B, f_j^I, f_j^I), j \in S_{of}^{max}$. For the case of fuzzy objective functions, the fuzzy goal is stated for each parameter of the triangular fuzzy number. Then the fuzzy goals are in the following form

$$\begin{aligned} \tilde{G}_j &= \left(\tilde{G}_j^l, \tilde{G}_j^m, \tilde{G}_j^u\right) = \left((f_j^{lI}, f_j^{lI}, f_j^{lB}), (f_j^{mI}, f_j^{mI}, f_j^{mB}), (f_j^{uI}, f_j^{uI}, f_j^{uB})\right) \qquad j \in F_{of}^{min} \\ \tilde{G}_j &= \left(\tilde{G}_j^l, \tilde{G}_j^m, \tilde{G}_j^u\right) = \left((f_j^{lB}, f_j^{lI}, f_j^{lI}), (f_j^{mB}, f_j^{mI}, f_j^{mI}), (f_j^{uB}, f_j^{uI}, f_j^{uI})\right) \qquad j \in F_{of}^{max} \end{aligned} . \tag{7}$$

Step 4 The mathematical model simultaneously finding appropriate extremes of the objective functions under specified conditions is formulated. This multi-objective model is transformed to the following one-objective form

$$\begin{aligned} &\max\ \alpha \\ &(1-w_j)\frac{f_j^B - f_j(\mathbf{x})}{f_j^B - f_j^I} \geq \alpha \qquad j \in S_{of}^{min}, \qquad (1-w_j)\frac{f_j(\mathbf{x}) - f_j^B}{f_j^I - f_j^B} \geq \alpha \qquad j \in S_{of}^{max} \\ &(1-w_j)\frac{f_j^{lB} - f_j^l(\mathbf{x})}{f_j^{lB} - f_j^{lI}} \geq \alpha, \qquad (1-w_j)\frac{f_j^{mB} - f_j^m(\mathbf{x})}{f_j^{mB} - f_j^{mI}} \geq \alpha, \qquad (1-w_j)\frac{f_j^{uB} - f_j^u(\mathbf{x})}{f_j^{uB} - f_j^{uI}} \geq \alpha \qquad j \in F_{of}^{min} \\ &(1-w_j)\frac{f_j^l(\mathbf{x}) - f_j^{lB}}{f_j^{lI} - f_j^{lB}} \geq \alpha, \qquad (1-w_j)\frac{f_j^m(\mathbf{x}) - f_j^{mB}}{f_j^{mI} - f_j^{mB}} \geq \alpha, \qquad (1-w_j)\frac{f_j^u(\mathbf{x}) - f_j^{uB}}{f_j^{uI} - f_j^{uB}} \geq \alpha \qquad j \in F_{of}^{max} \\ &\mathbf{x} \in X \\ &0 \leq \alpha \leq 1 \end{aligned} , \tag{8}$$

where α represents the weighted grade of membership of the solution, calculated as a minimum of values of all weighted membership function. Then, the real grade of membership of the solution is received by the elimination of the weights. The optimal solution can be formalized as a vector $\mathbf{x^o} = (x_1^o, x_2^o, ..., x_m^o)^T$ with the value of the *j*-th objective $f_j(\mathbf{x^o}), j \in S_{of}$, or $\tilde{f}_j(\mathbf{x^o}) = \left(f_j^l(\mathbf{x^o}), f_j^m(\mathbf{x^o}), f_j^u(\mathbf{x^o})\right), j \in F_{of}$.

If all functions are linear, the global extremes of the objective function in the models above can be found. In our investment practice, this fact is fulfilled. The variables in the models express the shares of assets in the portfolio. The objective functions, or their values, quantify the individual characteristics of the investment portfolio. Without 'special' investor preferences represented by additional constraints, the models are solvable by the simplex method. The fuzzy quantification of the characteristic may not be completely understandable to the investor. Then a simple or weighted form of averaging the three parameters of the triangular fuzzy number can be used. While the transformation to a single-valued indicator is more interpretable, some of the information about the dynamics of the process is lost.

3 Selecting a Portfolio with Open Unit Trusts

The second part of the article is devoted to a real-life portfolio selecting process with open unit trusts to demonstrate the application power of the proposed methodological concept. After specifying the investment policy, suitable investment instruments are selected. After downloading the data on the monitored characteristics, a mathematical model is formulated and then solved to select a portfolio of open unit trust. The results are analyzed.

3.1 Investment Policy, Selected Data

The aim of the paper is to provide a methodological framework for a selection of the investment portfolio (of open unit trusts) under conditions of unstable dynamics, which should be demonstrated in a case representing a significant part of investment practice. Also according to AKAT's annual press conference [1], the investment horizon is

being extended. Investors are focusing more on hedging in the distant future, especially in retirement. With longevity of investment goes hand in hand with less risk aversion, as evidenced by the increasing proportion in equity or real estate funds. In addition to retirement security, sustainability of investments is also an important theme.

In the light of the current market situation, long-term investment is chosen with the primary purpose of at least partial financial security in later retirement. The attitude to risk is moderate and slightly bold, particularly because of the longevity of the investment. Investment management is rather passive, without frequent reallocations, but with the possibility of regular investment. The steadily positive development of real estate funds in recent years calls for their presence in the portfolio, with a participation rate of at least ten percent. As the investor is not indifferent to the environment, the social climate or the management policy towards the employees, requires the participation of ESG funds in the same limiting proportion, meeting the environmental, social and governance criteria.

The three most important characteristics of open unit trusts - return, risk, cost - are monitored. Return is computed from the monthly prices according to (1)-(3). For the moving calculation, a one-month moving five-year period from January 2015 do December 2022 is used. Such a historical period has the potential to capture the long-term dynamics of the prices of open unit trust. The risk is determined by the absolute deviation of the calculated returns according to (4)-(5). Costs are represented by the annual charges associated with the investment in unit trusts (management, license, audit charges, etc.).

Fund name	**Return** [%]			**Risk** [%]	**Cost** [%]
Sporoinvest	-0.475	0.005	0.370	0.058	0.67
Sporobond	-1.899	-0.003	2.041	0.400	0.92
Trendbond	-4.888	-0.354	3.790	0.296	1.82
High Yield dluhopisový	-8.956	0.126	4.582	0.789	1.52
Korporátní dluhopisový	-5.963	0.078	3.831	0.255	1.40
Akciový Mix	-12.734	0.378	8.548	0.830	2.59
Dynamický Mix	-9.343	0.225	6.155	0.631	2.14
Vyvážený Mix	-6.706	0.172	4.446	0.464	1.59
Konzervativní Mix	-3.610	0.090	2.222	0.269	1.06
Fond řízených výnosů	-0.548	-0.075	0.352	0.100	1.60
ČS fond ŽC 2030	-7.696	0.223	5.145	1.967	2.36
ESG MIX 50	-6.954	0.112	4.261	0.525	2.15
ESG MIX 30	-4.336	0.077	3.127	0.427	1.51
ESG MIX 10	-3.245	0.041	1.894	0.253	1.35
Top Stocks	-23.270	0.705	19.146	2.217	2.31
Sporotrend	-17.419	0.278	13.600	1.993	2.35
Global Stocks	-9.310	0.500	9.361	0.360	2.66
REICO ČS Nemovitostní	-0.613	0.256	2.358	0.147	1.92

Table 1 Open unit trusts data, Source: Investment Centre [7], self-calculation

3.2 Portfolio Selection Model

To select a portfolio, the following mathematical model, following (8), is formulated as

$$
\begin{aligned}
&\max\ \alpha \\
&(1-0.45)\frac{\mathbf{r}_{\mathbf{min}}^{\mathbf{T}}\mathbf{x}-(-17.741)}{-1.043-(-17.741)} \geq \alpha, \qquad (1-0.45)\frac{\mathbf{r}_{\mathbf{ave}}^{\mathbf{T}}\mathbf{x}-0.037}{0.541-0.037} \geq \alpha, \qquad (1-0.45)\frac{\mathbf{r}_{\mathbf{max}}^{\mathbf{T}}\mathbf{x}-0.874}{14.49-0.874} \geq \alpha \\
&(1-0.4)\frac{1.671-\mathbf{d}^{\mathbf{T}}\mathbf{x}}{1.671-0.106} \geq \alpha, \qquad (1-0.15)\frac{2.239-\mathbf{n}^{\mathbf{T}}\mathbf{x}}{2.239-0.931} \geq \alpha \\
&\mathbf{e}^{\mathbf{T}}\mathbf{x}=1 \\
&x_{12}+x_{13}+x_{14} \geq 0.1 \\
&x_{18} \geq 0.1 \\
&0 \leq \alpha \leq 1
\end{aligned}
\qquad (9)
$$

,

where $\mathbf{x}=(x_1,x_2,...,x_{18})^T$ is a vector of variables expressing the share of the open unit trusts in order from Table 1, $\mathbf{r}_{\mathbf{min}}^{\mathbf{T}}=(r_1^{min},r_2^{min},...,r_{18}^{min})$, $\mathbf{r}_{\mathbf{ave}}^{\mathbf{T}}=(r_1^{ave},r_2^{ave},...,r_{18}^{ave})$ and $\mathbf{r}_{\mathbf{max}}^{\mathbf{T}}=(r_1^{max},r_2^{max},...,r_{18}^{max})$ are the vectors of return

parameters creating the triangular fuzzy number expressing the fuzzy return following (3), $\mathbf{d}^{\mathbf{T}} = (AD_1, AD_{2,}..., AD_{18})$ and $\mathbf{n}^{\mathbf{T}} = (co_1, co_{2,}..., co_{18})$ are the vectors including the risk (as absolute deviation) and costs associated with the particular open unit trust, $\mathbf{e}^{\mathbf{T}} = (1,1,...,1)$. The additional constraints ensure the required minimum share of the 'green', or real estate funds. The extreme values of the objectives are determined using the model (6) and similar. The weights are determined using scoring method according to the investor preferences in accordance with the investment policy specified above. Slightly higher score then gets the return before the risk. Even based on the author's investment experience, although not insignificant, the costs are not nearly as significant as the two criteria already mentioned.

3.3 Evaluation and Discussion over the Results

The linear model (9) is solved by Lingo optimization software. The portfolio has the following structure: 17.37% Sporobond, 20% ESG MIX 30, 32,55% Top Stocks, 0.83% Global Stocks and 29.25% REICO ČS Nemovitostní. The risk, or cost of the investment is estimated on 0.92%, or 1.8% level. The portfolio return is determined as triangular fuzzy number (-9.03, 0.32, 7.98)%. As can be noticed, for most funds, the difference of the lower and medium parameter of the fuzzy return is higher than that of the upper and medium one, indicating negative skewness of returns. The exceptions are the Sporobond and REICO ČS Nemovitostní. This fact, together with the second lowest risk level, determines a substantially higher share of real estate fund than the required minimum level. The unique 'positive' position of the minimum or maximum level of return relative to the mean value is indicative of Sporobond participation, although the intensity of the relationship is not nearly as significant as for a real estate fund. The aforementioned position of extreme values, combined with the level of risk, is also decisive in the selection of ESG funds that must be included in the portfolio. On the other hand, it can be seen that ESG funds are not the most interesting from a 'return-risk-cost' perspective, as they are included to the minimum extent possible, which nevertheless indicates other important portfolio characteristics for the investor. The biggest emphasis on returns gives the opportunity to participate in equity funds. An example is the significant participation of the Top Stocks fund.

The (fuzzy) targets are met at around 50% for the two most important criteria, of course slightly better for return than for risk. This means that the real value of the characteristics is approximately in the middle of the interval defined by the extremes. The weaker performance is not surprising for the less important costs. Although Sporoinvest has the highest level of stability, it has a relatively low expected return. The lowest level of costs can no longer prevent its absence from the portfolio. However, if risk were more accentuated, then Sporoinvest would start to participate in the portfolio. But more importantly, the stock component (partly at the expense of the Top Stocks, and REICO ČS Nemovitostní) would be boosted by Global Stocks, which exhibits relatively low risk.

Let us make a comparison with a portfolio that is based on a 'static' model in the sense of stable dynamics. Thus, consider the expected return just as an average indicator composed of moving averages (see r_i^{ave}), and the risk measured by the absolute deviation of the moving returns. The aforementioned positive position of the mean of returns relative to its extremes for the Sporobond ESG Mix 30 fund is not explicitly taken into account in the 'static' model. Thus, these funds are crowded out by funds with better level of risk, namely Sporoinvest (14.67%) and ESG Mix 10 (20%). The position of Top Stocks (22.41%) and REICO ČS Nemovitostní (42.92%) is virtually unchanged.

4 Conclusion

The article introduces the concept of unstable dynamics of prices, or returns of investment instruments, which should be taken into account when making an investment portfolio. To this end, a methodological procedure integrating a fuzzy approach operating with a triplet of moving characteristics is proposed. Thus, it is shown that the inclusion of the evolution of the dynamics of the price, and hence return behavior, as opposed to the more common 'static' form of dynamics in the existing literature, has a considerable impact on the outcome, i.e. the composition of the portfolio and its characteristics. This leads to a more complex data analysis that is appropriately used in the portfolio selection, in this case of the increasingly popular open unit trusts, as part of the optimization process.

The process of quantifying unstable dynamics opens the door to further research. Not only the selection of a partial risk measure, but also its aggregation calculation based on the weights of the parameters of the triangular fuzzy number are available for more detailed study. In particular, from the perspective of the role of extremes, the choice of the historical time period for the calculation of the moving characteristics should also be considered, so that in the general case it may not unduly bias the result.

Acknowledgements

The research project was supported by Grant No. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] AKAT ČR. (2023). *Výroční tisková konference AKAT*. [online]. Available at: https://www.akatcr.cz/Dokumenty/Aktuality/v253rocn237-tiskov225-zpr225va-akat-investice-do-fondu-v-cesk201-republice-prekrocily-hranici-jednoho-bilionu-kc [cited 2023-04-03]

[2] Ballestero, E., Günther, M., Pla-Santamaria, D. & Stummer, C. (2007). Portfolio selection under strict uncertainty: A multi-criteria methodology and its application to the Frankfurt and Vienna Stock Exchanges. *European Journal of Operational Research*, *181*, 1476–1487.

[3] Bauder, D., Bodnar, T., Parolya, N. & Schmid, W. (2021). Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, *21*, 221–242.

[4] Borovička, A. (2020a). Algorithmic improvements of the KSU-STEM method verified on a fund portfolio selection. *Information*, *11*, 21 p.

[5] Borovička, A. (2020b). New complex fuzzy multiple objective programming procedure for a portfolio making under uncertainty. *Applied Soft Computing*, *96*, 1–22.

[6] Huang, X. (2007). Portfolio selection with fuzzy returns. *Journal of Intelligent & Fuzzy Systems*, *18*, 383–390.

[7] Investiční centrum. (2023). *Fondy*. [online]. Available at: https://cz.products.erstegroup.com/Retail/cs/Produkty/Fondy/StruC3uA1nky/PuC5u99ehled/index.phtml [cited 2023-02-15]

[8] Konno, H. & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to the Tokyo stock market. *Management Science*, *37*, 519–531.

[9] Lai, Y. J. & Hwang, C. L. (1996). *Fuzzy multiple objective decision making: methods and applications*. Berlin: Springer.

[10] Markowitz, H. M. (1952). Portfolio selection. *Journal of Science*, *7*, 77–91.

An Analysis of Populist Attitudes Using SEM Models

Ing. Slávka Bozděchová[1]

Abstract. In this paper, I investigate the character of people who vote for a populist party. My goal is to understand their success during the recent elections. In the 2017 elections, the so-called populist parties achieved unexpected electoral success. In this article, I deal with the question of what factors (individual, contextual) influence the decision to vote for so-called populist parties. Using structural equation models popular in the social and behavioral sciences, I estimate causality between observed and latent variables. The latent variable is perceived in the work as a variable that contains populist attitudes.
I focus on two populist parties in this paper in particular: „Svoboda a přímá demokracie“ from Czech Republic and „Alternative für Deutschland“ from Germany. I show that regions with a low population density and a higher number of unemployed people have the greatest influence on attitudes towards populism. Support for populist attitudes is also associated with negative attitudes towards groups (migrants and minorities).

Keywords: populism, structural equation modeling, populist attitudes

JEL Classification: C51, C52
AMS Classification: 62J05, 91-11

1 Introduction

One of the main topics of political science and public debate over the past decade revolves around the ongoing crisis of democratic representation and the rise of populist parties as the main symptom of this crisis. The rise of populism can be identified approximately six to seven years back, in particular in light of the massive migration crisis in 2015-2016 and the subsequent social, legal and other problems that began to arise with the countries concerned and other EU countries. As a result of this, the political party „Svoboda a přímá demokracie“ (SPD) was formed in the Czech Republic and the „Alternative für Deutschland“ (AfD) in the German environment. Both sides can be described as right-wing or far-right. At the SPD it is possible to meet the membership of the far right and the far left in the Czech political environment. Both political groups are strongly populist, radical, nationalist or authoritarian.

However, it is interesting that very little attention has been paid to those who vote for these parties so far. Do populist voters have anything in common? The lack of scientific attention to this issue is remarkable. After all, in order to understand the recent achievements of populist parties coming from different ideological backgrounds, it is critical to know what their voters look like.

We already know a lot about those who vote for the radical right party (for example from a study by [13]), and more and more also about the supporters of the radical left party [9]. But if we want to find studies on what populist voters have in common, there is not much to find. One such study by Pauwels (2014) focuses on the question of what the voters who vote for Western European populist parties have in common. This study focuses on three countries: Belgium, Germany and the Netherlands. The conclusion of this study is that there is no single social-demographic group that supports populist parties.

In the subconscious we have that voters who vote for populist parties are mostly those who have lower socio-economic status, so-called “defeats of globalization”, who are radical and dissatisfied with politics at the national and European level [6]. A study by [7] proves that the voters of the populist radical right actually have low socio-economic positions and are eurosceptics [12].

Our hypotheses are in their concepts closest to the study from [10].

The first established hypothesis, therefore, solves the problem of the socio-economic status of voters and is established as follows:

Hypothesis 1 Voters of populist parties have in common that they will be from below socio-economic positions.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, nám. W. Churchilla 1938/4 130 67 Praha 3 - Zižkov, bozs00@vse.cz

That means, for example, they will come from the lower classes, will be unemployed or will have lower incomes.

Various studies have actually found the assumption that populist parties on the edge of the political spectrum tend to express eurosceptic attitudes, and that supporters of these parties (left and right) also tend to be more eurosceptical. Therefore I expect that:

Hypothesis 2 Voters of populist parties have in common that they are inclined to eurosceptic thinking.

2 Research Methodology

2.1 Date Description

The data for the research was taken from a survey conducted in the Czech Republic and Germany. The data file contains 3102 observations. This is cross data.

The model is defined:

- demographic individual variables: *age, gender, education* and *right-wing person*
- contextual variables: *migration balance, natural growth, population density, percentage of foreigners, share of atheists, household income, housing units, populist voting, unemployment rate,* and proportion of the population over the *age of 60*
- dependent variables: *populism*, this variable is modeled as a conjunctive variable, which is defined as the arithmetic average of the 5 questions related to populist views.
- latent variables: *euroscepticism*, includes 2 questions and *populist attitudes* contains 5 questions.

To ensure comparability, all contextual variables with unconditionally positive values were transformed into a logarithmic scale (except natural growth and migration saldo).

Selected variables in the model reflect existing explanations of the influence of contextual variables on the support of far-right parties [1] and on political dissatisfaction [11]. These include the unemployment rate and household income as a measure of economic conditions. We use the proportion of foreigners and atheists to derive ethnic origin in both countries. In addition, we also use demographic variables including *population growth*, *population density, migration,* etc. [4]. The model thus includes all available (and comparable) contextual information for both countries at their district level.

2.2 Methodology

The common formulation of SEM models (as I describe below) assumes that a structural model captures causal relationships exclusively between latent variables. In the application there are both latent characters, which I measure using a set of multiple questionnaire items, and directly measurable (manifest) variables, which we do not understand as an indicator for some latent character. As explained by Bollen and Noble (2011), these variables can be classified as part of a common model formulation by producing a "individual" latent variable (i.e. it has no other observed variable assigned to it) and the corresponding disturbance in the measurement model is "devalued" by attributing zero variance.

Independent variables For example, if we had a model with only *gender* and two latent i independent variables, the matrix $\mathbf{\Lambda_x}$ would look like this:

$$\mathbf{\Lambda_x} = \begin{array}{l} \\ euroscepticism_1 \\ euroscepticism_2 \\ populist\_attitudes_1 \\ populist\_attitudes_2 \\ populist\_attitudes_3 \\ populist\_attitudes_4 \\ gender \end{array} \begin{array}{c} \begin{array}{ccc} es & ps & Gender \end{array} \\ \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & \lambda_{32} & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array},$$

where latent variables are distinguished by the big letter at the beginning (*gender* vs. *Gender*). This then corresponds to the limit that $\mathrm{var}(\epsilon_7) = 0$. In this way, we have actually identified the measurable *gender* and the latent *Gender*.

Because in the model there are **more independent variables** than I define above. Then we can write these variables by marking a set of row indexes for the variable $populist\_attitudes_1, \ldots, populist\_attitudes_4$ as I_{og}, and similarly by introducing the set of line indexes I_{es} and $I_{without\ latent}$ as well as the column indices J_{es}, J_{og} and $J_{without\ latent}$, we can summarize:

$$[\mathbf{\Lambda_x}]_{ij} = \begin{cases} \lambda_{ij} & \text{if } i \in I_{es}, j \in J_{es} \text{ or } i \in I_{og}, j \in J_{og}, \\ 1 & \text{if } i \in I_{without\ latent}, j \in J_{without\ latent}, \\ 0 & \text{other.} \end{cases}$$

The spread limit can be defined as $[\mathbf{\Theta_x}]_{ii} = 0$ for $i \in I_{without\ latent}$.

Dependent variable in model For a dependent variable, the procedure is similar, if we had only one dependent variable, the entry would look like this: $\mathbf{\Lambda_y} = 1, \mathbf{\Theta_y} = 0$. In our case, we will encounter a model that contains more dependent variables, then the matrix $\mathbf{\Lambda_y}$ has a size of $p \times 1$, where p is the number of measured indicators of the "populism" variable, the matrix $\mathbf{\Lambda_y}$ looks like this:

$$\mathbf{\Lambda_y} = \begin{array}{c} \\ populism_1 \\ populism_2 \\ populism_3 \\ populism_4 \\ populism_5 \end{array} \begin{array}{c} populism \\ \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{pmatrix} \end{array}.$$

SEM models allow you to simultaneously estimate the relationships between observed and unobserved variables and the relations between unobserved variables. They also allow to include simultaneously conjunctive and categorical observed variables and latent variables [5].

In order to define structural models, we need to establish measurement models, which define relationship between the observed and the latent variable. The algebraically formulated model of measurement is presented in a system of two structural equations 1 and 2 written in a matrix shape.

$$x = \mathbf{\Lambda_x}\xi + \delta, \tag{1}$$

$$y = \mathbf{\Lambda_y}\eta + \epsilon, \tag{2}$$

In our case, the measurement model looks as follows:
$x = (euroscepticism\_1, euroscepticism\_2, populist\_attitudes\_1, populist\_attitudes\_2, 'populist\_attitudes\_3, populist\_attitudes\_4)$
$\delta = (es, ps)$

$$\mathbf{\Lambda_x} = \begin{bmatrix} \Lambda_{11} & 0 \\ \Lambda_{21} & 0 \\ 0 & \Lambda_{32} \\ 0 & \Lambda_{42} \\ 0 & \Lambda_{52} \\ 0 & \Lambda_{62} \end{bmatrix}$$

3 Applications

Results of the first model

All achieved results of statistical and diagnostic values were calculated using RStudio software. Most variables were not statistically significant at the level of significance 0.05. Among the statistically significant variables there are only 4, namely: *ps, right–wing person, lower–education* and *unemployment*.

From the table 1 we can analyze diagnostic indices. Within the recommended limits, only two indices came out: the RMSEA index with a value of 0.061 and the SRMR index with 0.067. The p-value for $\chi^2(114)$ rose below the significance level of $\alpha = 0.05$, so we can reject the zero hypothesis of H_0 and assume that the estimated model

does not match the data. This statistics is sensitive to the size of the file. For us, this means that the test result when using a larger number of observations (in our case 3102 observations) can be too sensitive. In practice, even if there was a very low (partial) correlation between two variables, between which there is no arrow in the diagram, this test would evaluate it as a significant deviation from the model. Therefore, mostly for estimated models with more observations, we reject the zero hypothesis of H_0. The second indicator is the relative χ^2, which minimizes this impact. This test came out too high at 10,992 over the recommended values. Indices CFI (0.785) and TLI (0.736) should have a value higher than the threshold of 0.95, but they do not approach that value. The Akaike and Bayes Information Criteria (AIC and BIC) values do not say anything on their own, as the values of these indices are used when comparing different models. The model has only two acceptable indices, RMSEA index (0.061) and index of absolute adjustment (0.067). The values of these indices are below the recommended value of 0.08, the lower the number the more accurate the estimate data.

	χ^2	df	p-value	calculation rel. χ^2	rel. χ^2	
model1	1253.13	114	.000	$^{1253.13}/_{114}$	10.992	
	CFI	TLI	AIC	BIC	RMSEA	SRMR
model1	0.785	.736	54429.047	54624.060	.061	.067

Table 1 Diagnostic index's values.

Modification of the first model

With the diagnostic index values of the estimated model with individual and contextual variables and with the latent dependent variable populistic, we are not satisfied because most diagnostic values did not go within the recommended values. We will adjust the model.

Modification of the index will be used modification indices, MI). For each variable in the model, the modification index is calculated together with the expected value of the parameter change. (epc). The Modification Index is an estimate of the value that indicates how much the value of the χ^2 statistics would change if the corresponding covariance was added to the model and this parameter was freely estimated [3]. Changes in the model should be made gradually. After each change, check the change in the χ^2 test and other indexes to see if the change really helped. Only changes that are theoretically reasonable in terms of models should be made. It starts with the highest value. Covariants are preferred first between the residues of items that fall into one factor, then the other variables.

Results of the modification model

Due to the high modification index, the following variables were eliminated after gradual adjustment: *right-wing person, age18–34, age60plus, lower–education, higher–education and natural growth*. There are 12 variables in the model.

	Estimate	Std. Err.	p-value
es	-0.031	0.040	.445
ps	0.755	0.066	.000
gender	0.043	0.025	.083
unemployment	-0.268	0.082	.001
migration–balance	0.004	0.003	.198
density–pop	0.039	0.019	.038
foreigners	-0.020	0.042	.633
atheists	0.067	0.056	.226
income	-0.002	0.002	.310
housing–units	0.035	0.031	.255
populist–voting	-0.174	0.068	.097
age60plus	-0.006	0.007	.380

Table 2 Values of regression coefficients for the modified model

	χ^2	df	p-value	calculation rel. χ^2	rel. χ^2	
model1	1253.130	140	.000	$^{1253.13}/_{114}$	10.992	
model1modif	248.313	47	.000	$^{248.313}/_{47}$	5.283	
	CFI	TLI	AIC	BIC	RMSEA	SRMR
model1	0.785	.736	54429.047	54624.060	.061	.067
model1modif	.929	.902	42400.270	42538.286	.038	.044

Table 3 Diagnostic index's values for the modified model.

Conclusions

The table 2 shows a list of regression coefficients for the remaining 12 variables. In the model we have 3 statistically significant coefficients (*ps, unemployment* and *density*).

From the Table 3 it can be seen that the model1modif has the closest results to the recommended values of diagnostic indices and thus the best match with the data. The only test for which we cannot accept a zero hypothesis is the χ^2 test. Next in order is the relative χ^2 test, which came out on the edge of the recommended value of a to 5.283. When we look at other indices, we find that CFI (0.929) and TLI (0.902) are closest to the threshold of 0.95. Indices RMSEA (0.038) s SRMR (0.044) have the lowest values from the models. When a zero value for these indices represents a perfect estimate. The values of u BIC and AIC were lower than in the model1.

4 Summary

The aim of the paper was to answer the question of what the voters of populist parties look like. And so understand their recent election achievements.

I found that many contextual factors are highly correlated. Regions suffering from depopulation are often marked by an ageing population structure, and other indicators of socioeconomic decline are often characteristic of these regions. This suggests that the context effects we observe, and some of the contextual effects observed by others, will likely be driven by (different) combinations of conditions.

It turned out that voters who vote for populist parties are not more likely to have lower incomes, come from lower classes, or have lower education. Moreover, euroscepticism has not proved to be a variable that unites voters from populist parties.

Support for populist attitudes in the Czech Republic and the Federal Republic of Germany is linked to higher unemployment and low population density. This indicates a demographic decline (low birth rate). Another factor that influences populist attitudes is a negative attitude towards groups (immigrants and minorities). Euroscepticism and age have no influence on populist attitudes.

When we summarize our results and apply hypotheses to them:

Hypothesis 1: Populist party voters have in common that they will be from below socio-economic positions.

Hypothesis 2: Populist voters have in common that they are inclined to eurosceptic thinking.

The model disproves the No. 1 hypothesis that voters who vote for populist parties come from lower socio-economic positions. The second hypothesis will be rejected because the variable unemployment rate, the unemployability rate or the household income is not significant in the model. As expected from previous studies, both hypotheses should be confirmed. But we failed to confirm it.

Acknowledgements

The work was supported by the Internal Grant Agency of Prague University of Economics and Business under Grant F4/24/2023.

References

[1] Arzheimer, K. (2009). Contextual factors and the extreme right vote in Western Europe, 1980-2002. *American Journal of Political Science*, 53(2), 259–275.

[2] Bollen Kenneth, A. & Noble, M. D. (2011) Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15639–15646.

[3] Byrne, B. M. (2010) *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.* Second Edition. New York: Taylor Francis Group, 177, 8.

[4] Dvořák, T., Zouhar, J. & Treib, O. (2022) Why do populists cluster regionally? Individual and contextual drivers of populist attitudes and vote choices in Germany and the Czech Republic. *Political Geography*, 0, 0.

[5] Hoyle Rick, H. (2012) *Handbook of Structural Equation Modeling.* Guilford Publication, ISBN 978-1-60623-077-0.

[6] Kriesi, H., Grande, E., Lachat, R., Dolezal, M. & Bornschier, S. (2008) *West European Politics in the Age of Globalization.* Cambridge: Cambridge University Press.

[7] Lubbers, M., Gijsberts, M. & Scheepers, P. (2002) Extreme right-wing voting in Western Europe. *European Journal of Political Research*, 41(3): 345-378.

[8] Pauwels, T. (2014) *Populism in Western Europe: Comparing Belgium, Germany and the Netherlands.* New York: Routledge.

[9] Ramiro, L. (2016) Support for radical left parties in Western Europe: social background, ideology and political orientations. *European Political Science Review*, 8(1): 1-–23.

[10] Rooduijn, M. (2017) What unites the voter bases of populist parties? Comparing the electorates of 15 populist parties. *European Political Science Review*, 10(3), 351–368.

[11] Salomo, K. (2019) The residential context as source of deprivation: Impacts on the local political culture. Evidence from the East German state Thuringia. *Political Geography*, 69(June 2018), 103—117.

[12] Werts, H., Scheepers, P. & Lubbers, M. (2013) Euro-scepticism and radical right-wing voting in Europe, 2002-2008: social cleavages, socio-political attitudes and contextual characteristics determining voting for the radical right. *European Union Politics*, 14(2): 183—205.

[13] Zhirkov, K. (2014) Nativist but not alienated: a comparative perspective on the radical right vote in Western Europe. *Party Politics*, 20(2): 286–296.

ANP Model for Suggestions on whether to Lead a Project Agile or Waterfall

Matej Brnka[1], Jan Rydval[2], Petra Pavlickova[3]

Abstract. In projects, a massive amount of money is often involved, and proper project management contributes to project success and efficient spending of money. Nowadays, two approaches to project management are commonly used and inherit the whole process from initial analysis, design, and implementation to delivery. In the traditional approach (waterfall), each phase follows the previous one, and the next phase cannot start before the previous is completed. It means the customer gets the final product at the end of the project. The agile approach is based on the principle of iterative time and incremental product, which must be functional. The customer gets functional parts of the product during the project and can provide feedback to the supplier.
The paper discusses selecting a project management style based on multi-criteria decision-making using the Analytic Network Process (ANP). The ANP model's structure and its clusters' composition come out on the international project management standard PRICNE 2 ® Agile. The composition of the clusters and the individual decision criteria correspond with the individual attributes of the project hexagon (project constraints like time, cost, scope, benefit, risk and quality), the attributes of the project team (size, location, communication, working style) and the characteristics of the project team members (T-shape, Pi-shape, I-shape, X-shape) and the overall project focus, from IT projects, through marketing to industrial projects. The model results in suggestions for the project manager on how to approach the management of the project or its parts.

Keywords: Agile, Analytic Network Process, PRINCE 2, Project, Project Management, Project Team

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

Leading a project means planning, delegating, monitoring and controlling to achieve project objectives within the time, cost, quality, scope, benefits and risk targets [1]. Nowadays, we use three approaches to manage a project (1) traditional project management, followed by a waterfall, (2) agile project management and (3) hybrid approaches [19]. In waterfall, the project is split into phases, when the next phase cannot begin before the previous is fully completed. The problem is that, for example, testing the product is practically the penultimate phase; the customer cannot provide feedback in time and gets almost the final product. Customers should define precise expected results at the beginning of the project, and work packages, responsibilities and deadlines are planned from kick-off to completion. The focus is on following a plan as precisely as possible [2], [3], [4]. In 2001, the agile manifesto introduced a new approach to managing a project as a reaction to the unsuitable older methods, mainly for IT projects [10]. Agile methods do not focus on the following plan. However, the project team develops a solution step by step (increment) and coordinates a result with the customer in a short cycle (interaction or sprint). In these methods, customers can provide a short goal at the beginning of the project, and the goal is particularised with each increment. It means that the customer gets functional parts during the project and provides feedback to a project team [5], [10], [19]. In summary, traditional project management assumes everything around the project is predictable, the scope is clear and well-understood, and the project can be planned in depth. Agile project management relies, among other things, on customer feedback and product adjustments, emphasising the delivery of parts of the project.

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, brnka@pef.czu.cz
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, rydval@pef.czu.cz
[3] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamycka 129,165 00, Praha - Suchdol, pavlickovap@pef.czu.cz

To decide if leading a project is in a waterfall or agile style, many factors must be considered. These factors can have varying preferences (weights), making the decision process complex. It is appropriate to use a multi-criteria decision method. In multi-criteria decision problems, uncertainty and vagueness can affect the decision-making process. The different stakeholders may have different priorities, objectives or interests. Another crucial issue is the significance of subjective preferences and individual stakeholder attitudes. We use mathematical processes to support such complicated decisions to avoid these problems.

One of them is the Analytic Network Process (ANP). ANP have been used in several sectors, like project management, strategic planning, and marketing research [14], [6]. ANP includes interdependencies and relations between elements from the same or different levels of the hierarchy of the decision-problem network [20]. ANP is a generalization of the analytic hierarchy process (for more information, see [16], [21]).

This paper discusses supporting a decision to select an appropriate project management style (waterfall or agile) based on a multi-criteria decision-making model using the ANP. In our pilot study, the model's clusters composition comes from the project management international standard PRINCE2 ® Agile.

2 Materials and Methods

2.1 Agile Methods

Agile methods have been created to uncover better ways of software development. According to Agile Manifesto [10], they are based on four principles: individuals and interactions over processes and tools, working software over documentation, customer collaboration over contract negotiation and responding to change according to plan. IT and non-IT companies and their projects are transforming to agile their teams and the whole structure [5]. There are many agile frameworks, for example, Scrum, Kanban, Extreme Programming etc.

2.2 Differences between the Waterfall and Agile Project Management

The project comes from the six aspects that need to be controlled and managed: time, cost, scope, quality, risk and benefits [11]. The main objective of a project manager is to deliver the product on time, with assigned costs, with an agreed scope and manage risk and quality. PRINCE2 ® Agile calls it Project hexagon [1] and defines which should be fixed or varied (flexed). In agile are, quality and scope typically flexed. On the other hand, in the waterfall, time and cost are flexed, as mentioned in **Figure 1**. Fixed variables can flex intolerances, and they would be adaptable to management by exception if these exceed.

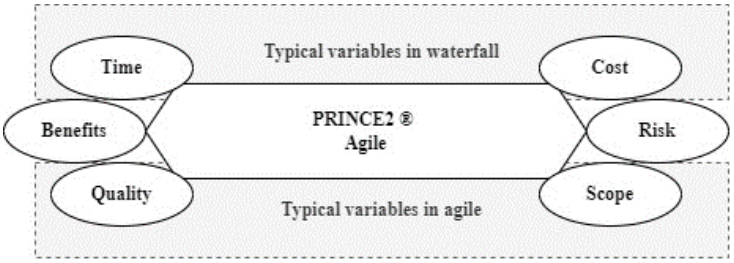


Figure 1 What to fix and what to flex, materials from the course, own processing

Another difference between waterfall and agile is delivery methods. As mentioned above, agile teams work iteratively with increments, to which customers should provide feedback. It means that the customer periodically, as part of the result, can test and decide the next direction of development. Agile frameworks, for example, Scrum, are more dedicated to smaller onsite teams to have frequent liaisons between members, and they should collaborate, cooperate and agree on the final result [1], [18]. PRINCE2 ® Agile recommends that team members are generalizing specialists (T – Shape) with detailed knowledge of one particular skill and less detailed knowledge of other skills [1]. If leading a project in a waterfall or agile style, it is necessary to use an appropriate decision approach to make a decision.

2.3 The Analytic Network Process

The Analytic Network Process (ANP) is a decision-making process that enables complex decisions for a variety of decision-making problems. The ANP as on pairwise comparisons of objects based procedure determine which element of each pair are favored. The ANP as a generalization of the Analytic Hierarchy Process (AHP), takes into account the various relationships between the elements at all levels of the decision network of the problem

structure hierarchy [20], i.e. the interdependencies between the elements of the decision structure. Many decision problems involve numbers of interactions and dependencies between decision elements (higher-level elements in a decision structure hierarchy on lower-level elements and between the same-level elements), and many decision problems cannot be therefore structured hierarchically, i.e., they cannot be structured into an AHP model (see [21] for an introduction to AHP theory). ANP is therefore more appropriate for complicated decision-making problems because it is represented by a network rather than a hierarchy [22], [20], [15].

Steps in the ANP process:

1. Decision-making problem structure describes dependency among decision elements. The ANP allows the inner dependence within a set of elements and the outer dependence among different sets of elements arranged in clusters.
2. Pairwise comparisons of the decision elements within and among the clusters of elements are performed and the pairwise comparisons matrix is created [20]. The consistency of these comparisons has to be controlled. The rate of consistency is measured by a consistency ratio CR defined by Saaty:

$$CR = \frac{CI}{RI_n} \tag{1}$$

where RI_n represents random inconsistency index. The *CI* stands for consistency index and for a comparisons matrix is calculated as:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{2}$$

where λ_{max} is the largest eigenvalue of Saaty's comparisons matrix and n is the number of criteria. A pairwise comparison matrix is considered inconsistent for CI>0.1, CI>0.15, respectively according to Saaty [17] and Muralidharan [12].

3. The priorities derived from the pairwise comparisons on the appropriate position form the Unweighted (original) Supermatrix [22], [21]. This Supermatrix has to be normalized using clusters weights, i.e. the Weighted Supermatrix is calculated.

$$\mathrm{W} = \begin{array}{c} \\ \mathrm{C_1} \\ \mathrm{C_2} \\ \vdots \\ \mathrm{C_N} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{C_1} & \mathrm{C_2} & \dots & \mathrm{C_N} \end{array} \\ \begin{bmatrix} \mathrm{W_{11}} & \mathrm{W_{12}} & \dots & \mathrm{W_{1n}} \\ \mathrm{W_{21}} & \mathrm{W_{22}} & \cdots & \mathrm{W_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{W_{n1}} & \mathrm{W_{n2}} & \dots & \mathrm{W_{nn}} \end{bmatrix} \end{array} \tag{3}$$

where each block ***Wij*** of the supermatrix consists of:

$$\mathrm{W_{ij}} = \begin{bmatrix} \mathrm{w_{11}} & \mathrm{w_{12}} & \dots & \mathrm{w_{1n}} \\ \mathrm{w_{21}} & \mathrm{w_{22}} & \cdots & \mathrm{w_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{w_{n1}} & \mathrm{w_{n2}} & \dots & \mathrm{w_{nn}} \end{bmatrix} \tag{4}$$

where:

$$\sum_{\mathrm{i}}^{\mathrm{n}} \mathrm{w_{ij}} = 1, \mathrm{j} \in \langle 1, \mathrm{n} \rangle \tag{5}$$

4. Stable overall preferences are obtained by raising the Weighted Supermatrix to powers, which generates the Limit Matrix. These preferences serve as the analysis (ranking respectively) of preferences of decision-making elements. See [20] for an introduction to the standard steps of the Limit Matrix calculation.

3 Results

We created a decision model to support the decision of which type of project management approach (waterfall or agile) to choose. We used the ANP method carried out by the SuperDecisions software [23]. We build a decision model as a relational network of individual decision elements grouped in defined clusters.

3.1 ANP Model Construction

Clusters' composition in the ANP model comes mainly from the differences between waterfall and agile project management approaches. The main objective of the constructed model is to find the preference of the project management style, whether to use a waterfall or agile approach to lead a project. In the model, we provide decision support for project management from the perspective of individual decision elements based on the pairwise comparison of various decision factors. The decision-maker will determine each cluster's and decision node's preferences, and the model output advises a suitable project management style.

Based on the project management definition, PRINCE2 ® Agile [1], we formed decision nodes, clusters and their relations in the ANP model (**Figure 2**). Detailed description of nodes, clusters and relations we introduced at

Model-Driven Organizational and Business Agility 2023 conference [4]. Our main objective is to decide whether to manage the project using waterfall or agile methodology. These two decision nodes are in the main cluster of Alternatives: Project Management Style. This main cluster is affected by 12 other clusters. Results of the ANP model provides the final suggestion to the project manager.

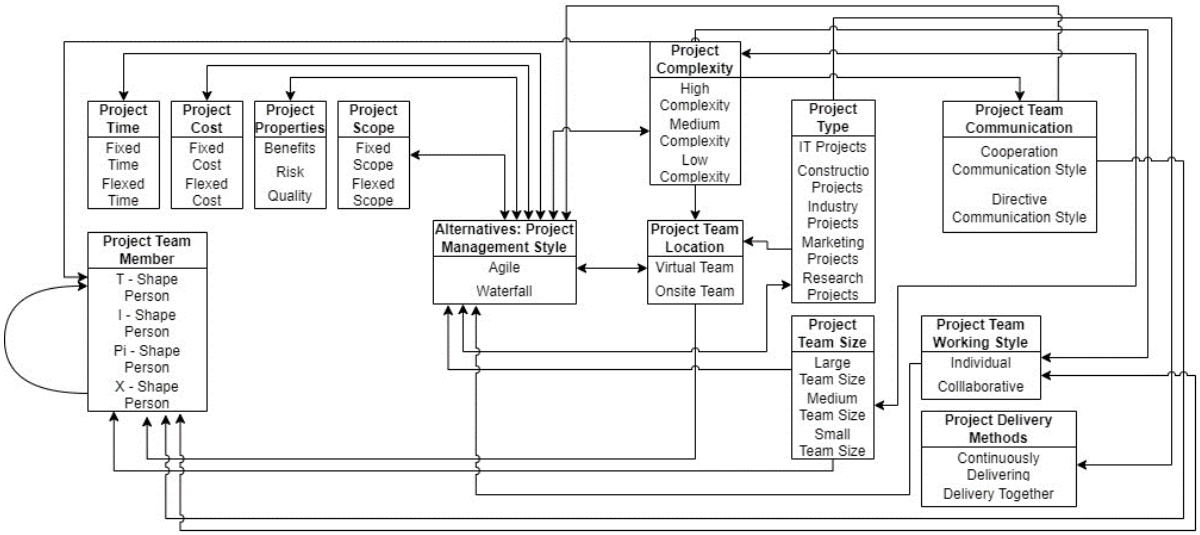


Figure 2 ANP Model (each block represents one cluster and its nodes)

Crucial to the success of the project is the project team. Using Agilometer methodology [1] we added clusters Project Members, Project Team Working Style and Project Team Communication to our model. Project Delivery Methods and Project Team Size come from Agile Manifesto and Scrum Guide [18], [10]. Based on our experience in project management, we identified clusters of Project Type and Complexity. Our model has five clusters (Project Team Member, Project Team Working Style, Project Team Communication, Project Team Size and Project Team Location) concerning project team communication and collaboration.

For example, a Project Team Member has four decision nodes: I – Shape, T – Shape, Pi – Shape, X – Shape, which represents the knowledge and skills of the team member. I – Shape people have a single and deep knowledge of the area. T – Shape people have deep knowledge of one area and superficial knowledge of another. The main difference between T – Shape and Pi – Shape people is that Pi – Shape people have two areas with deep knowledge. X – Shape people have unique human capabilities and can adapt well to the new environment. The next four clusters Project Properties, Project Scope, Project Time and Project Cost are related to the Project hexagon. Clusters: Project Cost, Project Scope and Project Time are based on the project triangle. Extension of project triangle to a project hexagon, according to PRINCE2 ® Agile [1], adds cluster Project Properties. Based on the principles of agile or waterfall, these variables can be fixed or flexed, depending on the project management approach. Fixed Project Scope means that it is given and planned what is to be done in the project. On the other hand, in flexed (various) Project Scope, we can change the project deliverables arbitrarily over time based on customer feedback. The last three clusters Project Type, Complexity and Delivery Methods contain characteristics of the project. Project outputs can be delivered to the customer together at the end of the project or continuously during the project. Each of the methods uses a different approach to the delivery of the output.

The individual relations (arrows in our model in **Figure 2**) represent the direction of influence of each cluster or decision node. For example, in terms of Project Management Style and its node Agile, we ask about the importance of fixed or flexed cost nodes, within the Project Cost cluster, and in the context of quantifying the model (pairwise comparison), the decision maker expresses this by his preferences.

3.2 Data for the ANP Model

The respondents' preferences for each decision element were obtained through pairwise comparisons using appropriately prepared questionnaire form (**Figure 3**). In the sample of our questionnaire form, the respondent is preferring the size of the project team in terms of agility. The first row contains the question: In terms of agile project management, do you prefer a large or medium size team? The second and third rows serve for comparing large and small team size and medium and small team size. The similar table forms are used also for other interesting factors (size, communication, cost etc.) and project management types (agile, waterfall).

However, in matrices with three or more decision elements, the respondent is struggling to maintain consistency in his decisions. In the case of inconsistent decisions, the inconsistency needs to be treated. This can be done through re-evaluation with the respondent in question, which is usually very challenging. Other ways of solving inconsistent decisions are discussed, for example, by Bláhová and Brožová [3] or Hlavatý and Brožová [8].

Comparisons with respect to Agile in Project Team Size

	9	7	5	3	1	3	5	7	9	
Large						x				Medium
Large									x	Small
Medium							x			Small

Figure 3 Pairwise comparison of Project Team Size with respect to Agile

3.3 Pilot Study – Selecting Suitable Project Management Style

These pilot study contains three respondents with experiences especially from IT projects managed by agile methodologies. The respondents determined preferences between decision nodes using pairwise comparison based on relations in ANP model. In this case the worst consistency ratio is CR = 0.074 which is less than 0.1, the limit by Saaty. The **Table 1** shows the average of the global preferences values of all nodes of the decision model. Cluster Project Management Style shows, that our model recommends managing the project more likely in an agile way (65.27 %) or at least to use agile methodologies based on the answers of the respondents.

Cluster	Decision node	Value	Cluster	Decision node	Value	Cluster	Decision node	Value
Project Management Style	Agile	0,2044	**Project Type**	Construction Projects	0,0063	**Project Team Communication**	Cooperation Communication Style	0,0193
	Waterfall	0,1088		Industry Projects	0,0035		Directive Communication Style	0,0112
Project Time	Fixed Time	0,0177		IT Projects	0,0095	**Project Working Style**	Collaboration Worikng Style	0,1133
	Flexed Time	0,0108		Marketing Projects	0,0046		Individual Worikng Style	0,0403
Project Cost	Fixed Cost	0,0177		Research Projects	0,0046	**Project Member**	I-shaped Persone	0,0570
	Flexed Cost	0,0108	**Project Team Size**	Large Size Team	0,0024		Pi-shaped Persone	0,0977
Project Scope	Fixed Scope	0,0108		Medium Size Team	0,0106		T-shaped Persone	0,0762
	Flexed Scope	0,0177		Small Size Team	0,0175		X-shaped Persone	0,0154
Project Properties	Benefits	0,0046	**Project Team Location**	Onsite Team	0,0248	**Project Complexity**	High Complexity	0,0013
	Quality	0,0082		Virtual Team	0,0113		Low Complexity	0,0055
	Risk	0,0156	**Delivery Method**	Continuously Delivering	0,0238		Medium Complexity	0,0034
				Delivery Together	0,0135			

Table 1 Limit Matrix of ANP model - raw values

The results show that fixed time, cost and flexed scope are important for agile project management, which corresponds to PRINCE2 ® Agile. Furthermore, the results show that the Pi-Shape personality is the most important for agile management, while the X-shape personality is not important for it, what corresponds to the team coach approach in the agile environment.

4 Discussion

Project management decision problems tend to be more complex; therefore, a mathematical tool is needed to account for this complexity. Therefore, the use of an ANP model is appropriate. It is used in complex decision-making environments by Ziemba [24], Bartoska [2], Havazik [6] and Rydval [14].

Our pilot study demonstrates the appropriateness of building an ANP model that includes all relevant decision nodes and clusters and their interrelationships. The pairwise comparison method was understandable and manageable for respondents, especially in comparing 2-3 decision elements within which the decision maker himself can also ensure consistency. When comparing four or more elements, it was difficult for the decision maker to maintain consistency, as confirmed by Hlavaty [7], and the comparison was very laborious, as confirmed by Havazik [6], Rydval [14]. Therefore, we need to find a better way of maintaining consistency, which is what Blahova and Brozova [3] or Hlavaty and Brozova [9] come up with.

5 Conclusion

The main goal of this article was to present a model to support the decision on how to run the project - agile or waterfall. Therefore, in this paper, we proposed the model and obtained data on the importance of its decision elements using ANP. In a pilot study, we demonstrated the model's suitability and reliability and obtained each decision node's preference values. The results obtained from IT industry experts recommend conducting projects more agile (65.27%) or at least using agile methodologies. In the next phase of the research, a questionnaire survey for pairwise comparison of the decision elements of the proposed model will be conducted with a wide range of project managers in different project management areas according to the type of project. This collected data will

be further used to develop an interactive application to support the project manager's decision on project management style.

Acknowledgements

The results and knowledge included herein have been obtained owing to support from the following institutional grant. Internal grant agency of the Faculty of Economics and Management, Czech University of Life Sciences Prague, grant no. 2022A0009.

References

[1] Axelos (2015). *PRINCE2 Agile*, First edition. Norwich: TSO, The Stationery Office.
[2] Bartoska, J., Jedlanova, T. & Rydval J. (2019). Semantic Model of Project Management in Corporate practice, in *37th International Conference on Mathematical Methods in Economics 2019*, (pp. 308–313).
[3] Blahova, P. & Brozova, H. (2022)., DSS Capabilities Evaluation Using ANP with Methods for Consistency Improvement, in *40th International Conference on Mathematical Methods in Economics 2022*, (pp. 21–26).
[4] Brnka, M., Pavlickova, P. & Rydval, J. (2023). ANP Model as a Support to Decide the Project Management Style, in *Model-driven Organizational and Business Agility 2023*.
[5] Ciric, D., Lalic, B., Gracanin, D., Tasic, N., Delic, M. & Medic, N. (2019). Agile vs. Traditional approach in project management: Strategies, challenges and reasons to introduce agile, in *Procedia Manufacturing*, Elsevier B.V., (pp. 1407–1414). https://doi.org/10.1016/j.promfg.2020.01.314.
[6] Havazik, O., Pavlickova, P. & Rydval, J. (2022). Model Design for Team Roles in Agile IT Projects, in *40th International Conference Mathematical Methods in Economics 2022*, Tolsteho 1556-16, Jihlava, 586 01, Czech Republic: Coll Polytechnics Jihlava, (pp. 91–97).
[7] Hlavaty, R. (2014). Saaty's matrix revisited: Securing the consistency of pairwise comparisons, in *Mathematical Methods in Economics 2014*, J. Talasova, (pp. 287–292).
[8] Hlavaty, R. & Brozova, H. (2022). Optimisation Approach to Dealing with Saaty's Inconsistency, in *40th International Conference on Mathematical Methods in Economics 2022*, (pp. 104–109).
[9] Hlavaty, R. & Brozova, H. (2022). Optimisation Approach to Dealing with Saaty's Inconsistency, in *40th International Conference Mathematical Methods in Economics 2022*, Tolsteho 1556-16, Jihlava, 586 01, Czech Republic: Coll Polytechnics Jihlava, (pp. 104–109).
[10] Kent, B. & Thomas, D. (2021). Manifesto for Agile Software Development.
[11] Kerzner, H. R. (2009). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. Wiley.
[12] Muralidharan, C., Anantharaman, N. & Deshmukh, S. G. (2023). Confidence Interval Approach to Consistency Ratio Rule in the Applications of Analytic Hierarchy Process, *West Indian Journal of Engineering*, vol. 26.
[13] Royce, W. W. (1970). *Managing the Develompent of Large Software Systems*.
[14] Rydval, J., Bartoska, J. & Jedlanova, T. (2019). Sensitivity Analysis of Priorities of Project Team Roles Using the ANP Model, in *37th International Conference Mathematical Methods in Economic2019*, Studentska 13, Ceske Budejovice, 37005, Czech Republic: UNIV South Bohemia Ceske Budejovice, (pp. 320–325).
[15] 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*.
[16] Saaty, T. L. (1986). *Axiomatic Foundation of the Analytic Hierarchy Process*.
[17] Saaty, T. L. (1977). *A Scaling Method for Priorities in Hierarchical Structures*.
[18] Schwaber, K. & Sutherland, J. (2020). *The Scrum Guide The Definitive Guide to Scrum: The Rules of the Game*.
[19] Thesing, T., Feldmann, C. & Burchardt, M. (2021). Agile versus Waterfall Project Management: Decision model for selecting the appropriate approach to a project, in *Procedia Computer Science*, Elsevier B.V., (pp. 746–756). https://doi.org/10.1016/j.procs.2021.01.227.
[20] Saaty, T. L. (2001). *Decision Making with Dependence and Feedback: The Analytic Network Process*. Pittsburgh: RWS Publications.
[21] Saaty, T.L. (2000). *Fundamentals of the Analytic Hierarchy Process*. Pittsburgh: RWS Publications.
[22] Saaty, T. L. (1996). *Decision Making with Dependence and Feedback: The Analytic Network Process*. RWS.
[23] Saaty, T. L. & Whitaker Saaty R. (1996). *SuperDecisions Software for Decision-Making*.
[24] Ziemba, P. (2019). Inter-Criteria Dependencies-Based Decision Support in the Sustainable wind Energy Management, *Energies (Basel) 12 (4)*, https://doi.org/10.3390/en12040749.

Mathematical Modeling of Ground Handling Process for Cargo Aircraft

Jakub Cíleček[1], Dušan Teichmann[2], Polina Yuryevna Koriukina[3], Cong Thanh Luu[4]

Abstract. Aircraft ground handling is a process involving a large number of sub-activities. Its basic task is to ensure the check-in of the aircraft after arrival and before departure. The ground handling of aircraft requires a whole range of resources. Most often, these are personnel and technical resources, the maintenance of which is financially very expensive. From an operational point of view, it is important that the ground handling of a specific aircraft takes place as efficiently as possible. The effectiveness of ground handling is quantified by its time-consuming nature. It is required that the ground handling takes as little time as possible. The presented article is devoted to the issue of managing the ground handling of cargo aircraft using network analysis methods in the conditions of Ostrava International Airport.

Keywords: ground handling, aircraft, cargo, project management, critical path method

JEL Classification: C69, C44
AMS Classification: 90C10

1 Introduction

The basic requirement for the operation of means of transport is that their downtime be minimal. The reason for this requirement is that the operators of these means of transport have significant funds tied up in them. A means of transport that is idle does not bring any financial benefits to its operator. However, some types of downtime are necessary. A case of necessary downtime of means of transport is downtime intended for unloading and loading shipments. This downtime cannot be avoided, as it results from the nature of the transport process.

The information given in the previous paragraph also applies to air transport. In addition to the benefits resulting from effective ground handling for air carriers, there are similar benefits for ground aviation infrastructure operators - airports. Effective handling is especially important for airports with a lower number of check-in stands or with a higher traffic intensity. Unnecessary downtime of aircraft on aprons directly negatively affects the capacity of the airport expressed in the number of handled flights for the pre-selected planning period. Another benefit for the airport can also be the maintenance of a lower number of technical means and personnel necessary to ensure the ground handling of arriving and departing aircraft.

When providing ground handling, it usually happens that air carrier´s requirements for the scope of ground handling will vary. Therefore, the ground handling of aircraft of different carriers can also include different numbers of activities.

Network analysis methods can be a tool for design or verification of ground handling. When using network analysis methods, it is important to know whether it will be necessary to control a process that is repeated or will be a one-time process. The critical path method can be used to model a process with fixed durations, the PERT assumes a process with stochastic durations. In the case of the presented article, the CPM method will be used, because the ground handling of the same type of aircraft of the same air carrier with the same contractually ensured range of ground handling will be modeled, which operates a regular airline line with the number of take-offs and landings three times a week to the international airport in Ostrava.

[1] VŠB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, jakub.cilecek@vsb.cz
[2] VŠB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, dusan.teichmann@vsb.cz
[3] VŠB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, polina.koriukina.st@vsb.cz
[4] VŠB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, cong.thanh.luu.st@vsb.cz

2 State of Art

The issue of optimizing the ground handling of aircraft at airports has already been addressed in the past. The articles are devoted to the optimization of ground handling in the conditions of specific airports, e.g. Amsterdam – Shiphol in work [3] or Košice [10]. Some authors, on the other hand, focus on the optimization of ground handling for aircraft of specific air carriers at hub airports [1].

In addition to the use of CPM, Gantt charts [7], the CDM (Collaborative Decision Making) method [6], heuristic approaches [2] or simulation methods, e.g. [12] and [11], are used in the literature for modeling ground handling. Some authors work with stochastic conditions [4], others also investigate other operational aspects in connection with ground handling. Other aspects include the issue of Air Traffic Management [9], business handling (passenger boarding strategy) and the effect of the seating arrangement in the aircraft cabin on the time of boarding the aircraft [8], the establishment of specialized infrastructure enabling faster transfers of handling equipment between aircraft stands [8] or the effect of operational emergencies, such as a strike by airport employees [5].

3 Critical Path Method – Short Summary of Used Method

The input data to Critical Path Method (CPM) is a list of sub-activities, the duration of sub-activities and information on the continuity of sub-activities. For each sub-activity, it must be known which activities immediately precede it and which activities follow it.

The input information is usually in tabular form, but it can be arranged in other ways.

If the number of activities is not too large, a so-called network graph $G[V;Y]$ is created, where V is the set of vertices of the graph and Y is the set of edges of the graph. Certain rules must be followed when creating a chart. A network graph has one start vertex (the vertex is labeled v_0) and one end vertex (the vertex is labeled v_n). These vertices represent the start and end of the project. The initial peak further represents the beginning of all activities that immediately follow the beginning of the project. The end peak represents the end of all activities that immediately precede the end of the project. The other vertices in the graph (vertices $v_1, \dots, v_{n-1}$) represent the beginnings and ends of sub-activities. Vertices are connected by oriented edges. An oriented edge $e \in Y$ usually represents a partial activity. An oriented edge representing a partial activity starts from a vertex representing its beginning, and at the same time representing the end of the immediately preceding activities. The oriented edge enters the vertex representing its end, and at the same time representing the beginning of the immediately following activities. When an activity representing waiting needs to be introduced into the project, a so-called dummy edge is introduced. Oriented edges are evaluated, and this evaluation represents the duration of individual sub-activities, in the case of fictitious edges representing waiting, then the waiting time. The evaluations of all oriented edges must be given in the same units. The graph must be acyclic.

The actual calculation takes place in the so-called internal calendar. We equate the expected real start time of the project with time 0 in the internal calendar. Subsequently, for each partial activity, we calculate their earliest possible start and end times. Consider two general network graph vertices $v_i \in V$ and $v_j \in V$, where $i \neq j$, which are connected by an edge originating from $v_i \in V$ and entering $v_j \in V$. Let us denote this activity by the symbol $[i;j]$, the evaluation of this edge y_{ij}, the time of the earliest possible start of this activity ES_{ij} and the time of the earliest possible end of this activity EF_{ij}. Applies to:

$$EF_{ij} = ES_{ij} + y_{ij} \tag{1}$$

The times of the earliest possible starts of the activity immediately following the start of the project are 0. The time of the earliest possible start of a sub-activity that is preceded by one sub-activity is equal to the time of the earliest possible end of the previous activity. When sub-activities $[i;j]$ are preceded by more activities (let's denote the set of sub-activities preceding sub-activity $[i;j]$ as $\bar{Y}_{[i;j]}$) then for the time of the earliest possible start of this sub-activity TE_i:

$$ES_{ij} = \max_{[s;i] \in \bar{Y}_{[i;j]}} \{EF_{si}\} = TE_i \tag{2}$$

If the times of the earliest possible ends of individual activities are calculated, then we calculate the end of the project T according to the relationship

$$T = \max_{[s;n]} \{EF_{sn}\} \tag{3}$$

In the next procedure, we calculate the latest possible beginnings and ends of individual sub-activities. We set the times of the latest possible ends of the activities immediately preceding the end of the project equal to T. Let us denote the time of the latest possible start of sub-activity $[i;j]$ as $LS_{ij}$and the time of the latest possible end of this activity LF_{ij}. Applies to:

$$LS_{ij} = LF_{ij} - y_{ij} \tag{4}$$

The time of the latest possible end of a sub-activity followed by one sub-activity is equal to the time of the latest possible start of the next activity. When sub-activity $[i;j]$ is followed by more sub-activities (let's denote the set of sub-activities following sub-activity $[i;j]$ as $\overline{\overline{Y}}_{[i;j]}$) then for the time of the latest possible end of this sub-activity

$$LF_{ij} = \min_{[j;r]\in\overline{\overline{Y}}_{[i;j]}} \{ES_{jr}\} = TL_j \tag{5}$$

The calculation on the network graph ends at the moment when the latest possible time of the activities immediately following the start of the project is determined.

Subsequently, four time reserves are calculated for each sub-activity $[i;j]$:

Total time reserve RC_{ij}

$$CR_{ij} = TL_{ij} - TE_i - y_{ij} \tag{6}$$

Free time reserve RV_{ij}

$$RV_{ij} = TE_j - TE_i - y_{ij} \tag{7}$$

Independent time reserve RN_{ij}

$$RN_{ij} = max\{0;\ TE_j - TL_i - y_{ij}\} \tag{8}$$

Dependent time reserve RZ_{ij}

$$RZ_{ij} = TL_j - TL_i - y_{ij} \tag{9}$$

Sub-activities with zero total time reserve represent the so-called critical path, which is the result of CPM and indicates critical activities. A critical activity is any activity whose delay or unplanned extension will cause a project delay. A critical path can also have several branches.

4 Application of Critical Path Method

The CPM method was used in modeling the ground handling of a cargo aircraft at Ostrava International Airport. The start of the project corresponds to the expected stopping time of the aircraft on the apron stand. The end of the project corresponds to the expected departure time of the aircraft from the apron stand. Table 1 gives a list of sub-activities performed during the ground handling of a cargo aircraft. The activity marked XX, which represents the connection of the Ground Power Unit (GPU), which is connected to the aircraft immediately at the beginning of the ground handling and disconnected immediately before its end. This activity takes place for almost the entire time of cargo aircraft check-in and will be one of the branches of the critical path. Table 1 also shows the values of individual types of reserves for individual sub-activities.

	Name of sub-activity	Length (min)	$[i;j]$
1	Marshalling	3	0;1
2	Securing of the wheels with chocks	2	1;2
3	Delineation of the protection zone	2	2;3
4	Connection of GPU	2	2;13
5	Visual check after arrival	3	3;15
6	Opening of cargo doors	2	1;4
7	Arrival of service vehicles	5	5;7
8	Unloading of containers from the main deck	30	7;9

9	Unloading of free loaded cargo space	15	7;8
10	Creation of loading instructions	3	4;5
11	Creation of loadsheet	3	5;6
12	Visual check by pilot	5	4;11
13	Loading of free loaded cargo space	15	8;10
14	Aircraft fueling	15	3;12
15	Loading of containers from the main deck	30	9;10
16	Finishing of weight and balance documents and handing over to crew	3	6;11
17	Departion of service vehicles	5	10;11
18	Closing of cargo doors	2	11;12
19	Visual check before departure	5	12;15
20	Cancellation of protection zone	2	15;16
21	Disconnection of GPU	2	14;15
22	Getting the wheel chocks off	3	16;17
23	Connection of push back vehicle	5	17;18
24	Pushing back the aircraft to taxiway	8	18;19
25	Disconnection of push back vehicle	3	19;20
26	Starting of aircraft engines	2	20;21
XX	GPU length of use	76	13;14

Table 1 Introductory information on the ground handling of a cargo aircraft at Ostrava International Airport

Critical activities are considered activities $[i; j]$, for which in Table 1, $RC_{ij} = 0$. The times of the earliest possible and latest possible beginnings and ends of individual sub-activities can be seen from the network graph, see Figure 1.

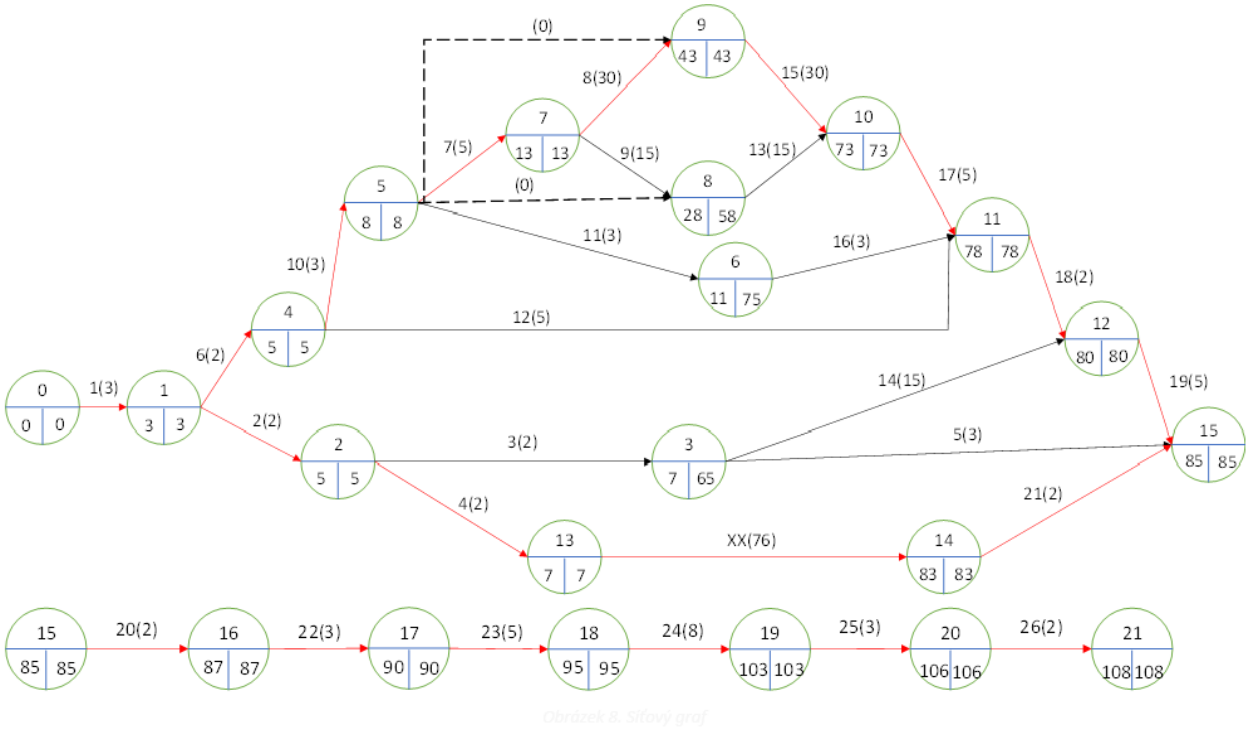


Figure 1 Network graph of ground handling of cargo aircraft at international airport Ostrava

The minimum ground handling time of the cargo aircraft was 108 minutes.
Computed time reserves are shown in Table 2.

	Name of sub-activity	$[i;j]$	y_{ij} $[min]$	RC_{ij}	RV_{ij}	RZ_{ij}	RN_{ij}
1	Marshalling	0;1	3	0	0	0	0
2	Securing of the wheels with chocks	1;2	2	0	0	0	0
3	Delineation of the protection zone	2;3	2	58	0	58	0
4	Connection of GPU	2;13	2	0	0	0	0
5	Visual check after arrival	3;15	3	75	75	17	17
6	Opening of cargo doors	1;4	2	0	0	0	0
7	Arrival of service vehicles	5;7	5	0	0	0	0
8	Unloading of containers from the main deck	7;9	30	0	0	0	0
9	Unloading of free loaded cargo space	7;8	15	30	0	30	0
10	Creation of loading instructions	4;5	3	0	0	0	0
11	Creation of loadsheet	5;6	3	64	0	64	0
12	Visual check by pilot	4;11	5	68	68	68	68
13	Loading of free loaded cargo space	8;10	15	30	30	0	0
14	Aircraft fueling	3;12	15	58	58	0	0
15	Loading of containers from the main deck	9;10	30	0	0	0	0
16	Finishing of weight and balance documents and handing over to crew	6;11	3	64	64	0	0
17	Departion of service vehicles	10;11	5	0	0	0	0
18	Closing of cargo doors	11;12	2	0	0	0	0
19	Visual check before departure	12;15	5	0	0	0	0
20	Cancellation of protection zone	15;16	2	0	0	0	0
21	Disconnection of GPU	14;15	2	0	0	0	0
22	Getting the wheel chocks off	16;17	3	0	0	0	0
23	Connection of push back vehicle	17;18	5	0	0	0	0
24	Pushing back the aircraft to taxiway	18;19	8	0	0	0	0
25	Disconnection of push back vehicle	19;20	3	0	0	0	0
26	Starting of aircraft engines	20;21	2	0	0	0	0
XX	GPU length of use	13;14	76	0	0	0	0

Table 2 Computed values of time reserves

5 Conclusion

In the presented article, a demonstration of the CPM method was shown in modeling the ground handling of a cargo aircraft. At the beginning of the article, some of the most important reasons for solving this type of task were mentioned. Modeled ground handling must be understood as a recurring project. Therefore, the CPM was then briefly described, and its practical use was shown on the example of a regularly recurring check-in of a specific type of aircraft and a specific carrier (whose business names had to be kept secret) at Ostrava International Airport. It has been proven that CPM has its practical application in these cases as well.

Acknowledgements

The creation of the article was supported by the VSB-Technical university of Ostrava internal grant project – TUO SP2023/087 Applied research, experimental development and innovation in transport and logistics.

References

[1] Al-Bazi, A., Ozturk, C. & Guimarans, D. (2016). Developing a Mathematical Model for Scheduling of Turnaround Operations (Low Cost Airline as a Case Study). *International Aviation Management Conference (IAMC)*. Dubai, United Arab Emirates, 16-25

[2] Andreatta, G., De Giovanni, L. & Monaci, M. (2014). A Fast Heuristic for Airport Ground-Service Equipment–and–Staff Allocation. *Procedia – Social and Behavioral Sciences.* 26-36. https://doi.org/10.1016/j.sbspro.2013.12.817.

[3] Beelaerts Van Blokland, W., Huijser, R., Stahls, R. & Santema, S. (2008). Future airport turnaround ground handling processes: How to reduce the turnaround time of aircraft at the airport. *Conference Proceedings of 10th TRAIL Congress*. Delft: TRAIL Research school, 1-25.

[4] Evler, J., Asadi, E. & Fricke, H. (2018). Stochastic Control of Turnarounds at HUB-Airports: A micro-scopic optimization model supporting recovery decisions in day-to-day airline ground operations. *8th SESAR Innovation Days Salzburg*, Austria

[5] Mantecchini, L., Malandri, C. & Reis, V. (2019). Aircraft turnaround and industrial actions: How ground handlers' strikes affect airport airside operational efficiency. *Journal of Air Transport Management.* 23-32. https://doi.org/10.1016/j.jairtraman.2019.04.007.

[6] Okwir, S., Ulfvengren, P., Angelis, J., Ruiz, F. & Nunez Guerrero, Y. M. (2016). Managing turnaround performance through Collaborative Decision Making. *Journal of Air Transport Management.* (58), 183-196. https://doi.org/10.1016/j.jairtraman.2016.10.008.

[7] Pavlas, R. (2007). Projektování systémů [Online]. Available at: http://www.elearn.vsb.cz/archivcd/FS/ProjS/Skripta_Projektovani_systemu.pdf [cited 2021-7-15].

[8] Schmidt, M. (2017). A review of aircraft turnaround operations and simulations. *Progress in Aerospace Sciences*. Germany, 2017, 25-38. https://doi.org/10.1016/j.paerosci.2017.05.002.

[9] Schultz, M. & Fricke, H. (2008). Improving Aircraft Turnaround Reliability. *ICRAT - International Conferences on Research in Air Transportation.*

[10] Strakova, E. & Mrva, P. (2011). Analysis the Utilization of Ground Support Equipment in Aircraft Ground Handling. In: *New trends in civil aviation 2011*. Prague, CTU in Prague. Publisher CTU in Prague. s. 113-116. ISBN 978-80-01-04893-1.

[11] Vidosavljevic, A. & Tosic, V. (2010). Modeling of Turnaround Process using Petri Nets. *World Conference of Air Transport Research Society.* Porto, Portugal.

[12] Wu, Ch. L. & Caves, R. E. (2004). Modelling and simulation of aircraft turnaround operations at air-ports. *Transportation Planning and Technology*. 25-46. https://doi.org/10.1080/0308106042000184445

Wavelet Method for Valuation of Options on Investment Project Expansion

Dana Černá [1]

Abstract. The paper addresses the pricing of options to expand, which are types of real options that allow expanding an investment project at a future date. Under the assumption that the underlying commodity price follows a geometric Brownian motion, the model for the valuation of the expansion option can be represented by a partial differential equation of the Black-Scholes type. Unlike options in finance, however, in the case of real options, there is no analytic solution, the payoff function itself is computed as a solution to a differential equation, and appropriate boundary functions have to be determined. The paper introduces the model for pricing expansion options and proposes a wavelet-based method for its numerical solution. The method employs the Crank-Nicolson scheme combined with the Galerkin method using a cubic spline wavelet basis. Numerical experiments are performed for the benchmark problem in the iron-ore mining industry. From a numerical point of view, the advantages are the higher-order convergence rate and the small number of iterations needed to resolve the problem with the required accuracy.

Keywords: real option, option to expand, cubic spline, wavelet, high-order convergence

JEL Classification: C44, G13
AMS Classification: 65M60, 65T60, 35Q91, 91G60

1 Introduction

In finance, an option gives the holder the right, but not the obligation, to buy or sell a financial asset at a specific price on a certain date. Real options apply this concept to investment project decisions involving risks and uncertainties, see [11, 12]. Incorporating real options analysis enables companies to make more informed decisions about the project, such as how much and when to invest in or abandon the project. Since the traditional discounted cash flow (DCF) approach has many limitations and cannot adequately capture the project flexibilities, several authors have suggested using partial differential equation (PDE) models. These models are similar to those for pricing options on financial assets, which have been widely used since the 1970s. Since the PDE approach takes into account not only time but also space dimensions, it allows for capturing spatial dependencies that the DCF approach cannot model. For a more detailed discussion of both approaches, we refer to [7, 11] and references therein.

This paper focuses on the valuation of expansion options, which are real options that allow for a specific price to expand an investment project in the future. We use a similar approach as in [1, 8, 9, 10] to obtain the PDE model for pricing expansion options under the assumption of the log-normal distribution of the underlying commodity price. The model is in some aspects similar to the standard Black-Scholes model but is more complicated because there is no analytic solution, the payoff function is not given but computed as the PDE solution, and boundary functions have to be appropriately determined.

We propose a method for the numerical solution of the PDEs representing the model, which combines the Galerkin method with modified cubic spline wavelets from [3] and the Crank-Nicolson scheme. The method is used to valuate an option to expand an investment project in the iron-ore mining industry with data taken from [10]. The numerical results confirm the validity of the method and show its advantages, which are higher-order convergence and the small number of iterations required to resolve the equation with the desired accuracy.

2 Model for Pricing Expansion Options

As in [6], the model assumes that the underlying commodity price P follows the process

$$dP = P\,(r - \delta)\,dt + P\sigma\,dW, \tag{1}$$

[1] Technical University of Liberec, Department of Mathematics, Studentská 2, 461 17 Liberec, Czech Republic, dana.cerna@tul.cz

where t denotes time, W is a Wiener process, σ is the volatility of P, r is the risk-free interest rate, and δ is the mean convenience yield on holding one unit of output.

We first introduce the quantities describing the model. Let $q_0(t)$ be the production rate for the project without any option, and $q_1(t)$ be the production rate for the project that increases the production by a factor κ at time T. Furthermore, $C(t)$ is the average cash cost rate of production per unit output, R is the rate of state royalties, and B is the company income tax rate. The after-tax cash flow rates are then given by

$$D_k(P,t) = q_k(t)\left(P(1-R) - C(t)\right)(1-B), \quad k = 0, 1, \tag{2}$$

see [9, 10]. Let $V_0(P,t)$ be the value of the project without any option, and $V_1(P,t)$ be the project with an embedded option enabling to extend the production by a factor κ at time T for the strike price K. It is known [1, 9, 10] that applying the contingent claim approach and Itô calculus, the fair value of the project $V_k(P,t)$, $k = 0, 1$, can be computed as the solution of a deterministic problem

$$\frac{\partial V_k}{\partial t} + \mathcal{L}V_k = -D_k, \quad \mathcal{L}V_k = \frac{P^2\sigma^2}{2}\frac{\partial^2 V_k}{\partial P^2} + (r-\delta)P\frac{\partial V_k}{\partial P} - rV_k, \tag{3}$$

where $P > 0$ and $t > T$. For numerical computation, it is convenient to replace the unbounded domain $(0,\infty)$ for P by a bounded domain $(0, P_{\max})$ with the maximal value $P_{\max}$ large enough.

Using equation (3), we find that in time $t > T$, the value $U(P,T)$ representing the difference $V_1(P,t) - V_0(P,t)$ is determined by

$$\frac{\partial U}{\partial t} + \mathcal{L}U = -D_1 + D_0, \quad P \in (0, P_{\max}), \quad t \in (T, T^*). \tag{4}$$

Time T^* is set to be the lifetime of the project, which implies that the terminal condition for equation (4) has the form

$$U(P,T^*) = 0, \quad P \in (0, P_{\max}). \tag{5}$$

For details concerning setting T^*, see Section 5. It remains to determine the boundary conditions. We employ the approach from [10], which is based on setting the net present value of the project for commodity prices $P = 0$ and $P = P_{\max}$ using

$$V_k(0,t) = \int_t^{T^*} D_k(0,\tau)\, e^{-r(\tau-t)}\, d\tau, \quad V_k(P_{\max},t) = \int_t^{T^*} D_k(P_{\max},\tau)\, e^{-r(\tau-t)}\, d\tau, \quad t \in (T, T^*). \tag{6}$$

This yields boundary conditions

$$U(0,t) = V_1(0,t) - V_0(0,t), \quad U(P_{\max},t) = V_1(P_{\max},t) - V_0(P_{\max},t), \quad t \in (T, T^*). \tag{7}$$

At time $t \in [0,T)$, the value of the expansion option $W(P,t)$ represents the increase in value $V_1(P,t)$ with respect to $V_0(P,t)$, and is therefore given by the backward equation

$$\frac{\partial W}{\partial t} + \mathcal{L}W = 0, \quad P \in (0, P_{\max}), \quad t \in (0,T). \tag{8}$$

As in [9], we equip the equation (8) with terminal and boundary conditions

$$\begin{aligned} W(P,T) &= \max\left(U(P,T) - K, 0\right), \quad P \in (0, P_{\max}), \\ W(0,t) &= 0, \quad W(P_{\max},t) = \max\left(U(P_{\max},T) - K, 0\right) e^{-r(T-t)}, \quad t \in (0,T). \end{aligned} \tag{9}$$

For another choice of boundary conditions, see [10]. To conclude, there are two PDEs to be solved. Firstly, function U is computed as the solution to the backward equation (4) with conditions (5) and (7), and then U is used to set conditions (9). Secondly, the function W representing the value of the expansion option is determined as the solution to the backward equation (8) with conditions (9).

3 Cubic Spline Wavelet Basis

The efficiency of the wavelet-based numerical method fundamentally depends on the wavelets used. Therefore, we propose a modification of the cubic spline wavelets on the unit interval from [4]. The resulting wavelets are then used in numerical experiments in Section 5, where it is shown that their use leads to an efficient method.

We start with the construction of so-called scaling functions. Let ϕ be a cubic B-spline given by

$$\phi(x) = \begin{cases} \frac{x^3}{6}, & x \in [0,1], \\ -\frac{x^3}{2} + 2x^2 - 2x + \frac{2}{3}, & x \in [1,2], \\ \frac{x^3}{2} - 4x^2 + 10x - \frac{22}{3}, & x \in [2,3], \\ \frac{(4-x)^3}{6}, & x \in [3,4], \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Furthermore, let ϕ_{b_1} and ϕ_{b_2} be boundary cubic B-splines defined as

$$\phi_{b_1}(x) = \begin{cases} \frac{7x^3}{4} - \frac{9x^2}{2} + 3x, & x \in [0,1], \\ \frac{(2-x)^3}{4}, & x \in [1,2], \\ 0, & \text{otherwise,} \end{cases} \qquad \phi_{b_2}(x) = \begin{cases} -\frac{11x^3}{12} + \frac{3x^2}{2}, & x \in [0,1], \\ \frac{7x^3}{12} - 3x^2 + \frac{9x}{2} - \frac{3}{2}, & x \in [1,2], \\ \frac{(3-x)^3}{6}, & x \in [2,3], \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Scaling functions $\phi_{j,k}$ are constructed as translations and dilations of these three generators. For $j \in \mathbb{N}$, $j \geq 2$, and $x \in [0,1]$,

$$\begin{aligned} \phi_{j,1}(x) &= 2^{j/2}\phi_{b_1}\left(2^j x\right), \quad \phi_{j,2}(x) = 2^{j/2}\phi_{b_2}\left(2^j x\right), \\ \phi_{j,k}(x) &= 2^{j/2}\phi\left(2^j x - k + 3\right), \; k = 3, \ldots 2^j - 1, \\ \phi_{j,2^j}(x) &= 2^{j/2}\phi_{b_2}\left(2^j - 2^j x\right), \quad \phi_{j,2^j+1}(x) = 2^{j/2}\phi_{b_1}\left(2^j - 2^j x\right). \end{aligned} \tag{12}$$

Wavelets are constructed in a similar way. First, a wavelet generator ψ is defined,

$$\psi(x) = \phi(2x-1) - 4\phi(2x-2) + 6\phi(2x-3) - 4\phi(2x-4) + \phi(2x-5). \tag{13}$$

Then, the wavelets in the inner part of the unit interval are constructed as

$$\psi_{j,k}(x) = 2^{j/2}\psi\left(2^j x - k + 3\right), \quad k = 3, ..., 2^j - 2, \quad j \in \mathbb{N}, \quad j \geq 2, \quad x \in [0,1]. \tag{14}$$

To construct boundary wavelets, we define wavelet generators as in [4],

$$\begin{aligned} \psi_{b1}(x) &= 6\phi_{b1}(2x) - \frac{57}{5}\phi_{b2}(2x) + \frac{919}{100}\phi(2x) - \frac{116}{25}\phi(2x-1) + \phi(2x-2) \\ \psi_{b2}(x) &= \frac{7}{3}\phi_{b2}(2x) - \frac{319}{60}\phi(2x) + \frac{101}{15}\phi(2x-1) - \frac{25}{6}\phi(2x-2) + \phi(2x-3). \end{aligned} \tag{15}$$

All the resulting cubic spline scaling and wavelet generators are displayed in Figure 1.

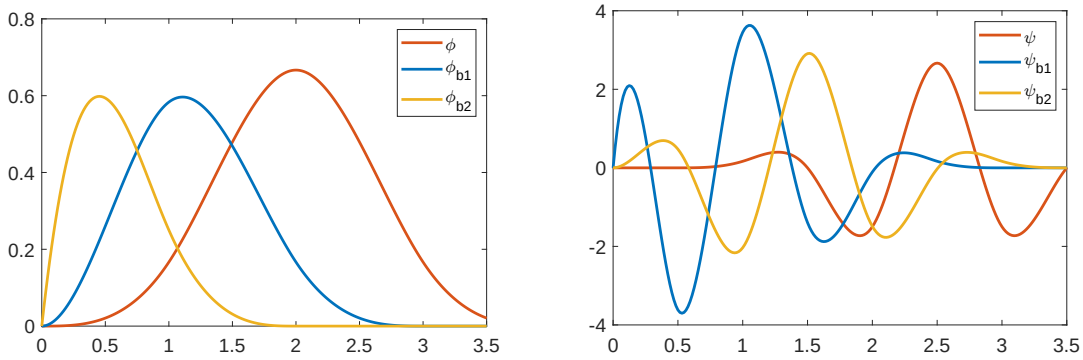


Figure 1 Cubic spline scaling and wavelet generators.

However, it is possible to take linear combinations of these boundary generators instead of taking these generators directly. The advantage is that the appropriate choice of coefficients of the linear combinations leads to better quantitative properties of the proposed method, namely a smaller number of iterations needed to solve a system of equations arising from discretization. Based on our numerical experiments, we define

$$\xi_{b1} = 0.90\phi_{b1} + 1.63\phi_{b2}, \quad \xi_{b2} = 0.69\phi_{b1} - 0.40\phi_{b2}. \tag{16}$$

For $j \in \mathbb{N}$, $j \geq 2$, and $x \in [0,1]$, the boundary wavelets are then generated as follows

$$\psi_{j,k}(x) = 2^{j/2}\xi_{b_k}\left(2^j x\right), \quad \psi_{j,2^j-k+1}(x) = 2^{j/2}\xi_{b_k}\left(2^j - 2^j x\right), \quad k = 1,2. \tag{17}$$

We now define the set Ψ^s comprising scaling functions on the coarsest level and s levels of wavelets functions,

$$\Psi^s = \Phi_2 \cup \bigcup_{j=2}^{2+s} \Psi_j, \quad \Phi_j = \left\{\phi_{j,k}, k = 1,\ldots,2^j+1\right\}, \quad \Psi_j = \left\{\psi_{j,k}, k = 1,\ldots,2^j\right\}. \tag{18}$$

Specifically, for $s = \infty$, the set Ψ^∞ is a wavelet basis of the space $L^2(0,1)$.

4 Wavelet Method

As presented in Section 2, the model is represented by two PDEs, which are both of the same type and can be written in the form

$$\frac{\partial u}{\partial t} + \mathcal{L}u = f(P,t), \quad P \in (0, P_{\max}), \quad t \in (T_1, T_2), \tag{19}$$

with terminal and boundary conditions

$$u(P,T_2) = g(P,T_2), \quad P \in (0,P_{\max}), \quad u(0,t) = h_0(t), \quad u(P_{\max},t) = h_1(t), \quad t \in (T_1,T_2). \tag{20}$$

In Section 3, the wavelet basis is constructed on the unit interval. To use the basis for the numerical solution, the basis can be either transformed to $(0, P_{\max})$ or the equation can be transformed to the interval $(0,1)$. Here, we choose the second approach and use the substitution $x = P/P_{\max}$. The next step is transformation to homogeneous Dirichlet boundary conditions. To this end, we define the function

$$w(x,t) = h_0(t) + (h_1(t) - h_0(t))\,x, \quad x \in [0,1], \quad t \in [T_1,T_2], \tag{21}$$

satisfying boundary conditions $w(0,t) = h_1(t)$ and $w(1,t) = h_2(t)$. The function $v(x,t) = u(xP_{\max},t) - w(x,t)$ is then the solution to the equation

$$\frac{\partial v}{\partial t} + \mathcal{L}v = f(xP_{\max},t) - \frac{\partial w}{\partial t} - \mathcal{L}w, \quad x \in (0,1), \quad t \in (T_1,T_2), \tag{22}$$

equipped with a terminal condition and homogeneous Dirichlet boundary conditions

$$v(x,T_2) = g(xP_{\max},T_2) - w(x,T_2), \quad x \in (0,1), \quad v(0,t) = v(1,t) = 0, \quad t \in (T_1,T_2). \tag{23}$$

Time discretization is performed using the Crank-Nicolson scheme. Let $M \in \mathbb{N}$ be a number of time steps and $\tau = (T_2 - T_1)/M$ be the step size. For $m = 0,\ldots,M$, denote $t_m = T_1 + m\tau$, $v_m(x) = v(x,t_m)$, and

$$f_m(x) = f(xP_{\max},t_m) - \frac{\partial w}{\partial t}(x,t_m) - \mathcal{L}w(x,t_m). \tag{24}$$

Since the equation to be solved is a backward equation, the scheme is also backward, and at each time instant t_m, $m = M-1,\ldots,0$, the task is to determine the function v_m using the scheme

$$\frac{v_{m+1} - v_m}{\tau} + \frac{\mathcal{L}(v_{m+1})}{2} + \frac{\mathcal{L}(v_m)}{2} = \frac{f_m + f_{m+1}}{2}. \tag{25}$$

Next, we apply the Galerkin method with the finite-dimensional spaces $V^s = \operatorname{span}\Psi^s$ generated by the wavelet bases Ψ^s from Section 3. The method consists in finding the function $v_m \in V^s$ such that

$$\frac{(v_m,z)}{\tau} - \frac{a(v_m,z)}{2} = \frac{a(v_{m+1},z)}{2} + \frac{(v_{m+1},z)}{\tau} - \frac{(f_m,z) + (f_{m+1},z)}{2}, \quad \forall\, z \in V^s, \tag{26}$$

where $(\cdot,\cdot)$ denotes the L^2–inner product and $a(u,z) = (\mathcal{L}(u),z)$. To transform this formulation to the system of linear equations, the solutions v_m is expanded in the basis Ψ^s, and the vector of coefficients of this expansion is denoted as $\mathbf{v}_m^s$,

$$v_m = \sum_{\psi_\lambda \in \Psi^s} \left(\mathbf{v}_m^s\right)_\lambda \psi_\lambda. \tag{27}$$

Using (27) and setting $z = \psi_\mu$. we obtain the system $\mathbf{A}^s\mathbf{v}_m^s = \mathbf{f}_m^s$ with

$$\mathbf{A}^s_{\mu,\lambda} = \frac{(\psi_\lambda,\psi_\mu)}{\tau} - \frac{a(\psi_\lambda,\psi_\mu)}{2}, \quad \left(\mathbf{f}_m^s\right)_\mu = \frac{a(v_{m+1},\psi_\mu)}{2} + \frac{(v_{m+1},\psi_\mu)}{\tau} - \frac{(f_m,\psi_\mu) + (f_{m+1},\psi_\mu)}{2}, \tag{28}$$

for indices λ and μ such that $\psi_\lambda, \psi_\mu \in \Psi^s$. Since the matrix $\mathbf{A}^s$ is non-symmetric, we solve the system by the generalized minimal residual method (GMRES) with Jacobi diagonal preconditioning.

5 Option to Expand an Iron Ore Mining Project

In this section, we present numerical experiments for the benchmark example from [9, 10] concerning an option to expand an investment project in the iron ore mining industry. The data of the project are the following:

$$Q = 10, \quad C(t) = 35e^{0.005t}, \quad B = 0.3, \quad R = 0.05, \quad r = 0.06, \quad \sigma = 0.2, \quad \delta = 0.02. \tag{29}$$

The production parameters are in billion tons. The production rate is $q_0(t) = 0.01\, Q\, e^{0.007t}$, and the strike price for the expansion at time $T = 2$ years is $K = 10$ billion US dollars. In the case of expansion, the production rate is multiplied by $\kappa = 2$. Thus, the production rate for the project with an embedded option on expansion is $q_1(t) = q_0(t)$ for $t \in [0, T)$ and $q_1(t) = \kappa\, q_0(t)$ for $t \geq T$. The lifetime T_0^* of the project V_0 and the lifetime T_1^* of the project V_1 are then computed from relations

$$Q = \int_0^{T_0^*} q_0(x)\, dx, \quad Q = \int_0^{T} q_0(x)\, dx + \kappa \int_T^{T_1^*} q_0(x)\, dx. \tag{30}$$

This yields $T^* = \max\left\{T_0^*, T_1^*\right\} = 75.804$ years. After the lifetime T_k^*, resources are already exhausted, and thus the production rate $q_k(t) = 0$ for $t > T_k^*$ for $k = 0, 1$. We set $P_{\max} = 100$ USD and compute the approximate solution using the proposed method. The resulting function representing the price of the option is depicted in Figure 2.

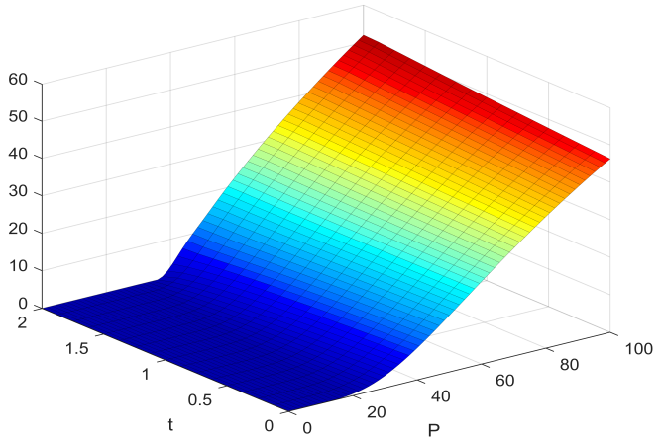


Figure 2 The function $W(P, t)$ representing the price of the option to expand (in billions of US dollars).

To study the efficiency of the method, quantities characterizing convergence are recorded in Table 1. The parameter s denotes the number of wavelet levels, N is the number of basis functions, M_1 is the number of time steps for equation (8), and M_2 is the number of time steps for equation (4). For the given level s and the corresponding parameters N, M_1, and M_1, we calculated option values $W_s = W(25, 0)$ (in billions of US dollars). Since the analytical solution is unknown, we cannot compute the resulting errors. Therefore, we use a similar approach as in [5] and compute differences $d_s = W_s - W_{s-1}$. These values characterize convergence in the sense that the error of order p implies that d_s is of order p. More precisely, if $h = C2^{-s}$ characterizes the spatial step and the error satisfies $e_s \sim O(h^p)$, then $d_s \sim O(h^p)$. Thus higher-order convergence of d_s suggests that the method is high-order convergent. The parameters it_1 and it_2 are the numbers of outer(inner) GMRES iterations in the last time step for equations (8) and (4) if the stopping criterion is that the relative residual is less than 10^{-12} and restart is after 10 iterations. We also tested the cubic spline wavelet basis from [2, 3]. While the errors were similar, the numbers of iterations were slightly higher than for the modification of the wavelet basis from [4] presented in Section 3, and therefore this paper only considers the basis from [4].

6 Conclusion

Valuation of real options using models represented by PDEs is a challenging and active field of research. Under the assumption of the log-normal distribution of the underlying commodity price, the model is similar to the famous Black-Scholes model. However, the difference is that the payoff function is not given but computed as the solution of a differential equation, and functions representing boundary conditions must be appropriately determined. The proposed wavelet-based method uses the Galerkin method with a modified cubic spline wavelet basis from [4]

s	N	M_1	M_2	W_s	d_s	it_1	it_2
0	5	64	2	1.381804		1(5)	1(5)
1	9	256	8	1.465432	8.36e-2	1(9)	1(9)
2	17	1024	32	1.492602	2.72e-2	4(6)	4(9)
3	33	4096	128	1.494879	2.28e-3	5(2)	6(1)
4	65	16384	512	1.495353	4.75e-4	5(4)	6(7)

Table 1 The resulting option prices W_s, differences d_s, and numbers of GMRES iterations it_1 and it_2.

in combination with the Crank-Nicolson scheme. Numerical experiments for an option to expand an investment project in the iron ore industry correspond to previous results from [9, 10] and thus confirm that the method is relevant. From the numerical point of view, the benefits of the method are high-order convergence and the fact that the method requires a relatively small number of iterations needed to resolve the problem with the desired accuracy. The future aim is to develop an efficient wavelet-based method for more complex models of real options pricing, such as models with stochastic volatility, multi-asset models, and models for other types of real options, such as options to defer and options to abandon.

Acknowledgements

This work was supported by grant No. GA22-17028S funded by the Czech Science Foundation.

References

[1] Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–659.

[2] Černá, D. & Finěk, V. (2014). Cubic spline wavelets with short support for fourth-order problems. *Applied Mathematics and Computation*, 243, pp. 44–56.

[3] Černá, D. & Finěk, V. (2015). Wavelet basis of cubic splines on the hypercube satisfying homogeneous boundary conditions. *International Journal of Wavelets, Multiresolution and Information Processing*, 13, 1550014.

[4] Černá, D. (2019). Cubic spline wavelets with four vanishing moments on the interval and their applications to option pricing under Kou model. *International Journal of Wavelets, Multiresolution and Information Processing*, 17, 1850061.

[5] Clift, S.S. & Peter A.F. (2008). Numerical solution of two asset jump diffusion models for option valuation. *Applied Numerical Mathematics* , 58, pp. 743–782.

[6] Cortazar, G. & Schwartz E.S. & Casassus, J. (2001). Optimal exploration investments under price and geological-technical uncertainty: a real options model. *R & D Management*, 31, 181–189.

[7] Dixit, A. & Pindyck, R. (1994). *Investment Under Uncertainty.* Princeton: Princeton University Press.

[8] Haque, M. & Topal, E. & Lilford, E. (2014). A numerical study for a mining project using real options valuation under commodity price uncertainty. *Resources Policy*, 39, 115–123.

[9] Hozman, J. & Tichý, T. (2021). Numerical valuation of the investment project with expansion options based on the PDE approach. In R. Hlavatý (Ed.) *39th International Conference on Mathematical Methods in Economics* (pp. 185-190). Prague: Czech University of Life Sciences.

[10] Li, N. & Wang, S. (2019). Pricing options on investment project expansions under commodity price uncertainty. *Journal of Industrial & Management Optimization* , 15, 261–273.

[11] Myers, S.C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5, 147–175.

[12] Mun, J. (2002). *Real Options Analysis: Tools and Techniques for Valuing Strategic Investments and Decisions*. Hoboken: John Wiley & Sons.

Statistical Analysis of Brand Marketing

Andrea Čížků[1]

Abstract. The goal of the paper is to investigate brand marketing strategy for different generations of consumers. Brand marketing strategies aims to be more personalized as different types of customers perceive the brand value differently. Specifically, generational approach is applied and sources of brand value are analyzed across different generations of customers. Questionnaire survey data focused on identification of sources of brand value across different generations are statistically analyzed by applying chi-squared tests within contingency tables. The results show that (1) generational aspects play an important role in brand marketing strategy as the perceived source of the brand value turned out to be dependent on consumers' age, (2) this dependence is quite stable over time.

Keywords: generational stratification, brand value, contingency table, chi-squared test

JEL Classification: C12, D12, M30, M31, M37
AMS Classification: 91B82

1 Introduction

Brand of a product is one of the main factors influencing success of a company as consumers can choose from a large number of alternative products which makes their decision-making process difficult. The brand is a symbol of quality and thus plays an important role in consumers' buying behavior. The effective brand marketing strategy must always be tailored to target specific segments of consumers. The focus of this paper is on different generations of consumers and their perceptions of the brand value sources as it is supposed that there might be significant diversity brand value sources perceptions across different generations.

The question of generational stratification in the context of identifying sources of brand value is widely discussed in the literature. Gajanová et al. [1] perform statistical analysis across generational cohorts in banking industry. Kisieliauskas, Jančaitis [2] study an impact of green marketing on perceived brand value in different generations. Majerová et al. [3] analyze the theory of generational stratification in the context of brand marketing communication strategy and discuss practical implications. Slabá [4] focuses especially on age as a significant factor influencing consumer buying behavior.

The question of how sources of brand value are dependent on different generations of consumers is statistically tested in this paper. Specifically, the aim of the statistical analysis is to answer the following questions:

1) Are there differences in perceiving sources of the brand value between different generations?
2) Are these differences between generations stable in time?

The structure of the paper is as follows. Chapter 2 describes data and statistical methodology. Results are presented and discussed in chapter 3. The final chapter 4 summarizes main conclusions.

2 Data and methodology

Two questionnaire surveys among Slovak respondents were conducted in 2018 and 2019 and performed by Majerová et al. [3]. Each survey was carried out on a statistical sample of 1978 respondents. The questionnaire investigated general perception of brand value sources as well as demographic characteristics of respondents. Generations were defined as follows:
baby boomers: born in 1946–1964;
generation X: born in 1965–1976;
generation Y: born in 1977–1994;
generation Z: born in 1995–2010.

[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.

Statistical analysis of the data from questionnaire is based on contingency tables. The formulated hypotheses are statistically tested by applying chi-squared tests. Specifically, the chi-squared test of independence between two categorical variables is applied. Chi-square statistic is calculated as follows:

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}\right] \tag{1}$$

where O_{ij} is the observed frequency in the *i*-th row and *j*-th column of a contingency table,

E_{ij} is the expected frequency in the *i*-th row and *j*-th column of a contingency table,

m is the number of rows and n represents the number of columns in a contingency table.

Expected frequencies are in contingency tables calculated according to the formula:

$$E_{ij} = \frac{O_{i\bullet} \cdot O_{\bullet j}}{N} \tag{2}$$

where $O_{i\bullet} = \sum_{j=1}^{n} O_{ij}$ is the sum of observed frequencies in the *i*-th row,

$O_{\bullet j} = \sum_{i=1}^{m} O_{ij}$ is the sum of observed frequencies in the *j*-th column,

$N = \sum_{i=1}^{m}\sum_{j=1}^{n} O_{ij}$ is the total number of observations.

If the null hypothesis H_0 : 'Categorical variables are independent' is true, then the chi-square statistic χ^2 has a chi-square distribution with $k = (m-1)\cdot(n-1)$ degrees of freedom, i.e. $\chi^2 \sim \chi^2(k)$. The critical region is defined as $\left(\chi^2_{1-\alpha}(k);\infty\right)$, where $\chi^2_{1-\alpha}(k)$ is called a critical value and is obtained as $(1-\alpha)$-quantile of the distribution $\chi^2(k)$, α being the level of significance ($\alpha = 0{,}05$). The null hypothesis is rejected if $\chi^2 \in \left(\chi^2_{1-\alpha}(k);\infty\right)$, otherwise it is not rejected.

3 Results and discussion

There are a great variety of different aspects of the brand value and its sources. Technically, answers to each question from each questionnaire (performed in 2018 and 2019) are expressed as a categorical variable (or a set of categorical variables). Each categorical variable is then combined together with the variable generation to form a contingency table within which the chi-squared test of independence is performed to test the formulated hypotheses. Questions in the questionnaire are grouped into four main categories:

- questions regarding preferences and brand-name characteristics,
- questions regarding consumers' feelings,
- questions regarding respondents' attitudes,
- questions regarding customers' expectations,
- questions regarding respondents' utility.

Questions from all these five categories are discussed in detail in the following subchapters.

3.1 Questions regarding preferences and brand-name characteristics

Formulation of the questions in the questionnaire:

1) Mark the extent to which you agree/disagree with the following statements:
   a. I prefer brand-name products.
   b. I consider brand-name products to be superior in quality.
   c. Brand-name products provide me with a prestige and image which is important to me.

Strongly disagree				Strongly agree
1	2	3	4	5

2) What is the characteristic with the highest influence on your choice of the brand:
   a. product (quality, function)
   b. price (discounts, competitive prices)
   c. distribution (availability)
   d. communication (advertisements, image)
3) What is the reason that would induce you to change your favorite brand-name:
   a. quality
   b. price
   c. availability
   d. image

Three categorical variables were obtained for the three statements presented in the first question. Other two categorical variables for the remaining two questions. Each of them was combined with the variable generation to obtain the corresponding contingency table. For illustration, the first contingency table is as follows:

Generation	**I prefer brand-name products**					**Sum**
	1	**2**	**3**	**4**	**5**	
After the war generation	38	89	193	100	18	**438**
Generation X	39	83	215	92	34	**463**
Generation Y	64	154	259	167	61	**705**
Generation Z	45	97	109	98	23	**372**
Sum	**186**	**423**	**776**	**457**	**136**	**1978**

Table 1 Contingency table for variables Generation and Preference of brand-name products (based on questionnaire from 2018).

Data in this contingency table represent observed frequencies O_{ij}. The sums in the final row $O_{i\bullet}$ and column $O_{\bullet j}$ were used to calculate expected frequencies E_{ij} according to the formula (2). Chi-square test statistic χ^2 was then obtained from (1).
The same calculations were performed for the remaining categorical variables and results are summarized in the following table. The first value is the calculated χ^2 - statistic, the second is P-value and symbols *, ** and *** denote that the result is significant at 10%, 5% and 1% level of significance, respectively.

Statistic	**2018**					**2019**				
	Question					**Question**				
	1a)	**1b)**	**1c)**	**2**	**3**	**1a)**	**1b)**	**1c)**	**2**	**3**
χ^2	42,01	11,42	22,08	11,47	42,93	41,84	37,34	24,49	20,40	45,35
P-value	0,000	0,493	0,037	0,245	0,000	0,000	0,000	0,017	0,016	0,000
significance	***		**		***	***	***	**	**	***

Table 2 Results of the chi-square test for questions 1-3.

The null Hypothesis H_0 : 'Answers to given question don't depend on generation.' is rejected in all cases except the questions 1b) and 2) in the project from 2018. Therefore, the way respondents answered to these questions depended on the generation in most cases.

3.2 Questions regarding respondents' feelings

Formulation of the questions in the questionnaire:

4) Shopping brand-name products cause me to feel:

	Strongly disagree				**Strongly agree**
a) Prestige	1	2	3	4	5
b) Happiness	1	2	3	4	5
c) Enthusiasm	1	2	3	4	5
d) Expectations	1	2	3	4	5
e) Satisfaction	1	2	3	4	5
f) Blames	1	2	3	4	5
g) Confidence	1	2	3	4	5
h) Modern	1	2	3	4	5
i) Positive associations	1	2	3	4	5

χ^2 P-value significance	**Question**								
Year	**4a)**	**4b)**	**4c)**	**4d)**	**4e)**	**4f)**	**4g)**	**4h)**	**4i)**
2018	35,09 0,000 ***	13,67 0,323	34,19 0,001 ***	36,57 0,000 ***	15,23 0,229	12,25 0,426	40,23 0,000 ***	23,30 0,025 **	28,20 0,005 ***
2019	33,22 0,001 ***	24,34 0,018 **	35,94 0,000 ***	20,62 0,056 *	23,25 0,026 **	23,42 0,024 **	24,54 0,017 **	25,25 0,014 **	19,46 0,078 *

Table 3 Results of the chi-square test for the items in section 4 in the questionnaire.

The results show that the way how respondents answer to questions 4a)-4i) is mostly dependent on generation and that the dependence is seen more often in 2019 than in 2018.

3.3 Questions regarding respondents' attitudes

Formulation of the questions in the questionnaire:

5) My attitude towards brand-name products is:

	Strongly disagree				Strongly agree
a) I intentionally buy brand-name products.	1	2	3	4	5
b) I am interested in brand-name products regularly.	1	2	3	4	5
c) Brand-name products catch my attention as I consider them superior in quality.	1	2	3	4	5
d) Brand-name products catch my attention as I consider them more prestigious.	1	2	3	4	5
e) I am interested in brand name products only rarely.	1	2	3	4	5

Statistic	**2018**					**2019**				
	Question					**Question**				
	5a)	**5b)**	**5c)**	**5d)**	**5e)**	**5a)**	**5b)**	**5c)**	**5d)**	**5e)**
χ^2 P-value significance	16,68 0,162	24,40 0,018 **	24,32 0,018 **	24,26 0,019 **	31,38 0,002 ***	46,31 0,000 ***	50,54 0,000 ***	25,75 0,012 **	20,18 0,063 *	16,90 0,154

Table 4 Results of the chi-square test for questions 5a-5e.

The results clearly show that the answers to these questions about respondents' attitudes are dependent on a generation in most cases.

3.4 Questions regarding respondents' expectations

Formulation of the questions in the questionnaire:

6) I expect brand-name products to be:

	Strongly disagree				Strongly agree
a) popular	1	2	3	4	5
b) available	1	2	3	4	5
c) modern	1	2	3	4	5
d) superior	1	2	3	4	5
e) innovative	1	2	3	4	5
f) improving my image	1	2	3	4	5
g) good quality to price ratio	1	2	3	4	5
h) catching my attention	1	2	3	4	5
i) supporting national economy	1	2	3	4	5

The results of the performed chi-square test are summarized in the following table.

χ^2 P-value significance	**Question**								
Year	**6a)**	**6b)**	**6c)**	**6d)**	**6e)**	**6f)**	**6g)**	**6h)**	**6i)**
2018	33,37 0,001 ***	18,19 0,110	28,83 0,004 ***	14,42 0,274	22,70 0,030 **	21,66 0,042 **	21,66 0,042 **	30,90 0,002 ***	13,63 0,325
2019	36,76 0,000 ***	22,71 0,030 **	38,41 0,000 ***	21,20 0,047 **	13,49 0,335	22,30 0,034 **	28,29 0,005 ***	15,97 0,192	22,97 0,028 **

Table 5 Results of the chi-square test for the items in 6 in the questionnaire.

The results show that the answers to questions about respondents' expectations are mostly dependent on the generation.

3.5 Questions regarding respondents' utility

Formulation of the questions in the questionnaire:

7) My utilities from brand-name products are:

	Strongly disagree				Strongly agree
a) Happiness	1	2	3	4	5
b) Higher social status	1	2	3	4	5
c) Facilitated process of making new friends	1	2	3	4	5
d) Catches attention of others	1	2	3	4	5
e) Belongs to my life style	1	2	3	4	5

Results from the performed chi-square test of independence of answers to these questions on generation is summarized in the following table.

Statistic	**2018**					**2019**				
	Question					**Question**				
	7a)	**7b)**	**7c)**	**7d)**	**7e)**	**7a)**	**7b)**	**7c)**	**7d)**	**7e)**
χ^2 P-value significance	34,19 0,001 ***	13,88 0,308	25,85 0,011 **	12,71 0,390	28,87 0,004 ***	40,96 0,000 ***	35,90 0,000 ***	31,69 0,002 ***	28,50 0,005 ***	52,65 0,000 ***

Table 6 Results of the chi-square test for questions 1-3.

The again results show that the answers to these questions about respondents' utility from brand-name products are dependent on generation in most cases.

4 Conclusion

The paper presented main results from a large statistical survey focused on the analysis of the brand value sources with respect to different generations. To this end, chi-squared tests within contingency tables were applied to statistically test hypotheses about the dependence of the perceived brand value sources on different generations. The results show that the sources of the brand value are mostly dependent on generations. Moreover, this dependence proved to be quite stable over time in most cases. Similar results were obtained also in other empirical studies (Gajanová et al. [1]; Majerová et al. [3]; Slabá [4]). This result has significant implications for brand building as it is necessary to focus and target brand value communication strategy on individual generations in which people have similar perceptions of brand value sources.

References

[1] Gajanová, L., Nadanyiová, M., Majerová, J. & Aljarah, A. (2021). Brand Value Sources in Banking Industry: Evidence for Marketing Communication Across Generational Cohorts. *Polish Journal of Management Studies*, 23(1), 151-171. https://doi.org/10.17512/pjms.2021.23.1.10

[2] Kisieliauskas, J. & Jančaitis, A. (2022). Green Marketing Impact on Perceived Brand Value in Different Generations. *Management Theory and Studies for Rural Business and Infrastructure Development*, 44(2), 125-133. https://doi.org/10.15544/mts.2022.13

[3] Majerová, J., Čížků, A., Gajanová, L. & Nadanyiová, M. (2022). The theory of Generational Stratification in the Context of Brand Marketing Communication Strategy. *Intellectual Economics*, 16(2), 76-94. https://doi.org/10.13165/IE-22-16-2-05

[4] Slabá, M. (2019). The Impact of Age on the Customers Buying Behaviour and Attitude to Price. *Littera Scripta*, 12(2), 1-14. https://doi.org/10.36708/Littera_Scripta2019/2/11

Computational Aspects of Data Envelopment Analysis

Martin Dlouhý[1]

Abstract. The data envelopment analysis is a well-known non-parametric method of efficiency evaluation. The large number of production units in data envelopment analysis leads to extensive computations because one linear program has to be solved for each production unit. The objective of this paper is to summarise the current body of knowledge on computational aspects of data envelopment analysis. We can distinguish three main groups of approaches. Firstly, the preprocessing methods that can quickly classify or score a production unit without solving a linear program. Secondly, the linear programming methods aim to reduce the computational requirements of solving linear programs by accelerating their performance or extracting opportunistic information for classification from them. Thirdly, the big data methods deal with the computational challenges when large sets of production units are present. The computational experiments have shown that the savings achieved by applying some procedures can be significant.

Keywords: data envelopment analysis, convex hull.

JEL Classification: C61
AMS Classification: 90C05

1 Introduction

Data envelopment analysis (DEA) is a well-known non-parametric method of the relative efficiency evaluation. In the literature, we can find thousands of applications from various application fields [18]. DEA was developed by Charnes, Cooper, and Rhodes [7] in 1978 to assess the relative efficiency of a set of homogenous production units with multiple inputs and outputs. Charnes, Cooper, and Rhodes formulated linear programming (LP) model that classifies production units as efficient or inefficient and determines the level of relative efficiency. One disadvantage of this LP method is that DEA needs to solve one LP model for each production unit.

The objective of this paper is to review existing approaches to the reduction of the computational complexity of DEA. In the following sections, we summarise the current body of knowledge on the computational aspects of DEA. We can distinguish three main groups of methods: preprocessing, linear programming methods, and big data methods. The rest of the paper is organised as follows: Section 2 introduces the basic DEA model. Section 3 presents the preprocessing methods, Section 4 presents the LP methods, and Section 5 presents the big data methods. Section 6 is a conclusion.

2 Data Envelopment Analysis

DEA is a non-parametric method that does not need any assumptions about a functional form of the production frontier and no parameter estimation. DEA uses mathematical programming to construct the production (efficient) frontier as the piecewise linear envelopment of the observed data [7]. The method assumes that the production units use a clearly defined set of inputs to produce a defined set of outputs. On the other hand, the weights (prices) of inputs and outputs are unknown, and without information on input and output prices, the economic efficiency of production units cannot be calculated. DEA can calculate the technical efficiency of a production unit, which is calculated as the best possible ratio of the weighted output to the weighted input or vice versa.

In the multiplier form of the DEA model, each production unit optimises its input and output weights to maximise the technical efficiency score. DEA classifies the production units as technically efficient or inefficient depending on their location with respect to a production frontier. Technically efficient production units are those that lie on the production frontier.

The theory distinguishes the output-oriented DEA model, which maximises quantities of outputs produced by the fixed levels of inputs, and the input-oriented DEA model, which minimises quantities of inputs required to produce

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, dlouhy@vse.cz.

the fixed levels of output. According to the character of the returns to scale, there are four standard DEA models. The CCR model (Charnes, Cooper, and Rhodes) [7], in which constant returns to scale (CRS) are assumed. The BCC model (Banker, Charnes, and Cooper) [4], in which variable returns to scale (VRS) are assumed. The DEA models with non-decreasing returns to scale (NDRS) and non-increasing returns to scale (NIRS) were also developed. Many other DEA models have been developed since the formulation of the original DEA models [4][7], see for example [8],[10],[13].

The DEA models have two equivalent formulations: the multiplier form and the envelopment form. Let us suppose that we have a set of n homogenous production units that use m types of inputs to produce r types of outputs. The envelopment formulation of the input-oriented variable-returns-to-scale DEA model (1) for production unit q is below:

$$\begin{aligned} &minimise\ \theta_q \\ &subject\ to \\ &\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, \qquad i = 1,2,\dots,m, \\ &\sum_{j=1}^{n} y_{kj}\lambda_j \ge y_{kq}, \qquad k = 1,2,\dots,r, \\ &\lambda_j \ge 0, \qquad j = 1,2,\dots,n, \\ &\sum_{j=1}^{n} \lambda_j = 1, \end{aligned} \tag{1}$$

where θ_q is the technical efficiency score of production unit q, x_{ij} is the quantity of input i used by production unit j, y_{kj} is the quantity of output k produced by production unit j, λ_j is the intensity variable that measures the individual contribution of production unit j in the formation of the efficient target for production unit q. In the input-oriented model, the technical efficiency score θ_q represents a size of input reduction that makes production unit q technically efficient.

The envelopment formulation of the output-oriented variable-returns-to-scale DEA model (1) for production unit q is below:

$$\begin{aligned} &maximise\ \theta_q \\ &subject\ to \\ &\sum_{j=1}^{n} x_{ij}\lambda_j \le x_{iq}, \qquad i = 1,2,\dots,m, \\ &\sum_{j=1}^{n} y_{kj}\lambda_j \ge \theta_q y_{kq}, \qquad k = 1,2,\dots,r, \\ &\lambda_j \ge 0, \qquad j = 1,2,\dots,n, \\ &\sum_{j=1}^{n} \lambda_j = 1, \end{aligned}$$

In the output-oriented model, the technical efficiency score θ_q is greater or equal to one, representing a size of input expansion that makes production unit q technically efficient. Depending on the choice of a model orientation, a sequence of n linear programming models (1) or (2) has to be solved. Hence, DEA is quite a computationally intensive method because one LP model has to be solved for each production unit.

3 Preprocessing

Dulá and López [17] define a *preprocessor* as a procedure that can quickly classify or score a production unit without solving an LP model. This definition excludes procedures that reduce computational requirements, which somehow involve solving LP models either by accelerating their performance or extracting from them opportunistic information for classification. Preprocessors are not expected to classify or score all the production

units conclusively, but they are trying: (1) to reduce the total number of LP models that will eventually have to be solved and/or (2) to reduce their size so that they can be solved faster. Preprocessors can be classified as (a) procedures that identify efficient and extreme-efficient production units and (b) procedures that identify inefficient production units.

- *Sorting*. The production units with the minimum value of input *i* or maximum value of output *k* in the VRS or ADD DEA models are extreme-efficient if unique. If not unique, they correspond to production units on the boundary, which means they may or may not be efficient (e.g., weak efficient). These production units are identified by sorting the dimensions [1]. Maximally *m*+*r* extreme-efficient units can be identified this way. Let us have a set of six production units (Figure 1) and assume an input-oriented variable-returns-to-scale DEA model (1). According to the sorting procedure, unit (1;2) with minimal input and unit (6;7) with maximal output are extreme-efficient.
- *Norm maximisation*. Dulá, Helgason, and Hickman [16] propose identifying extreme production units of the convex hull by finding the element in the set of production units that maximises the Euclidean distance to an arbitrary point *p* in $\Re^{m+r}$. One such point can be constructed as a "worst virtual unit". In an illustrative example (Figure 1), the worst virtual unit is (6;2). Units (1;2) and (6;7) are identified as efficient units because they maximise the Euclidean distance to point (6;2).
- *Domination*. The idea of this procedure is quite simple: a dominated production unit is inefficient [2]. Sueyoshi and Chang [21] and Dulá and Lopez [17] observed that domination is a powerful preprocessor with the potential to classify a large proportion of inefficient units at a reasonably low cost. The procedure applies to all return-to-scale assumptions. In fact, this is an application of the Free Disposable Hull (FDH) [11] in which the evaluation of a production unit is based only on existing production units, not on their convex combinations. As a result, the production frontier is non-convex and does not require any prior assumption about the returns to scale. In an illustrative example, units (3;3) and (5;4) are dominated and inefficient.
- *Hyperplane translation*. This procedure is based on the fact that for any combination of input/output weights, there is at least one efficient unit [17]. The *output-input ratio analysis* suggested by Chen and Ali [9] shows that production units with the highest performance by ratio analysis are efficient. This ratio analysis is a special case of hyperplane translation with binary input/output weights (0/1). Production units (1;2) and (2;4) maximise output/input ratio and must be efficient.

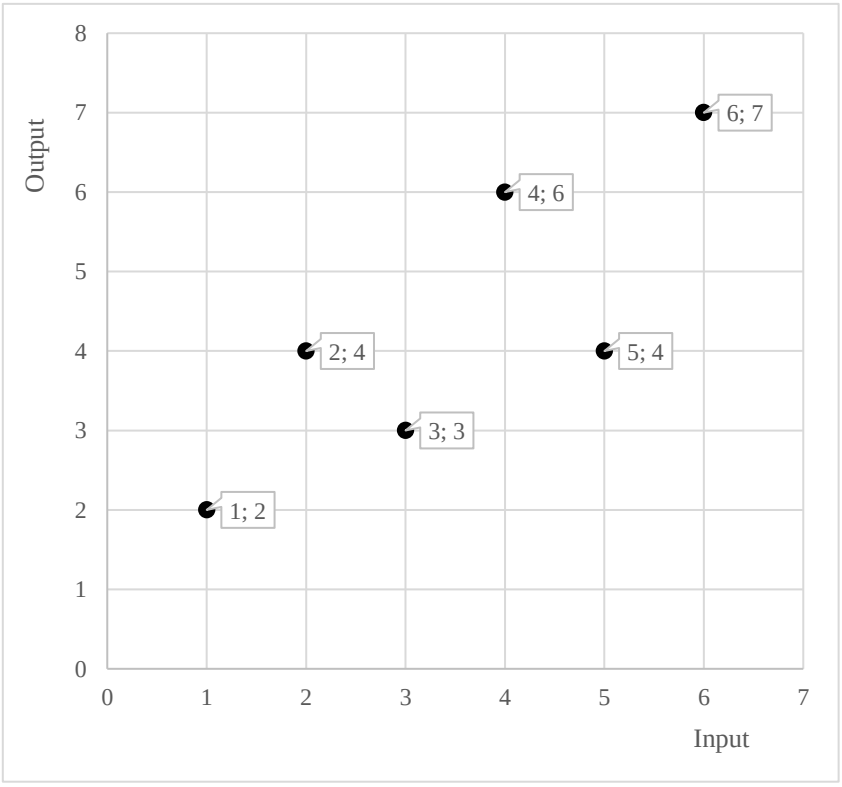


Figure 1 Illustrative example

4 Linear Programming Methods

The methods based on linear programming reduce the computational requirements of solving LP models either by accelerating their performance or extracting from them opportunistic information for classification. Some methods of this type are listed below:

- In any optimal solution of the DEA program, production units in the reference set for any other production unit have to be efficient [1][2]. Such information saves LP calculations for the previously identified efficient production units. This procedure is known as EIE (*Early Identification of Efficients*). The weakness of this procedure is that the proportion of efficient units in the set of units is usually low.
- If a production unit is classified as inefficient, it can be omitted from the LP models for remaining units. The effect is that the number of variables in the envelopment DEA models is progressively reduced. This procedure is known as RBE (*Reduced Basis Entry*).
- The *super-efficiency DEA model* of Andersen and Petersen [3] excludes the production unit under evaluation from the reference set in each envelopment model (also known as DD – Deleted Domain) and, at the same time, gives us additional information on super-efficiency for efficient units. The weakness is that the super-efficiency model with variable returns to scale can be infeasible for some efficient units. Although the model was not developed in order to simplify computations, however, it does.
- Dulá [15] developed the linear programming method *BuildHull*. In the first phase, the *BuildHull* procedure identifies the elements of the frame (the extreme elements of the polyhedral hull); in the second phase, the procedure scores the inefficient production units. The idea is that savings in computations achieved by using only the frame to score the inefficient units dramatically compensate for the effort identifying it. In the first phase, one initial efficient unit is determined, for example, the unit with the largest output in the VRS model, and then n-1 LP models is solved, which start with two units and gradually increases by one unit (variable) every time a new efficient unit (frame element) is found until the status of all units is determined. In the second phase, the standard DEA models for inefficient units are solved in which the number of variables is the cardinality of the frame plus one. Hence, the procedure is output-sensitive because the computational performance depends on the number of units in the frame. The *BuildHull* procedure is flexible in the sense that once a frame is identified, it can be applied to any of the four standard returns-to-scale DEA models (CRS, VRS, NIRS, NDRS) and to any model orientation.

5 Big Data Methods

The methods in this section are based on linear programming (Section 4), but they specifically address the computational challenges when large sets of production units are present. Both presented big data methods (hierarchical decomposition and single decomposition scheme) aim to determine the set of efficient production units as fast as possible in the first phase. In the second phase, the rest of the production units is scored.

- Barr and Durchholz [5] in 1997 proposed the method of *hierarchical decomposition* (HD) in which the set of production units is partitioned into smaller subsets (blocks). The inefficient production units are excluded, and the efficient production units are merged into a larger subset and evaluated again. The procedure is repeated until all globally efficient units are determined. Finally, the globally efficient units are used to score inefficient units in LP models with a smaller number of variables. HD was one of the first methods designed specifically for large-scale DEA computations.
- The framework of Khezrimotlagh et al. [19], also known as the *single decomposition scheme* (SD), differs from HD in that instead of partitioning the set of production units into several subsets and evaluating each subset, only a subsample of production units (one subset) is used to evaluate all other units. The authors recommend setting the cardinality of the subset at $\sqrt{n}$ and using heuristics with the goal of maximising the number of efficient production units in the initial subset. The inefficient (interior) production units can be excluded from the subset, while the exterior units are added to the subset. If the number of production units is very large, the units can be partitioned into several blocks, as in HD. According to Khezrimotlagh and Zhu [20], the SD framework outperforms the *BuildHull* method and hierarchical decomposition. For the first time in the DEA literature, they evaluated one million production units in their experiments.
- Dellnitz [12] improved the HD method and showed that his decomposition method outperforms the single decomposition framework of Khezrimotlagh et al. [19].

6 Conclusion

DEA is a well-known of the efficiency evaluation. It is a computationally intensive method because we have to solve one LP model for each production unit. The computational complexity of DEA can be a practical problem if the number of production units is extremely large. In this paper, we reviewed the methods and procedures that deal with the computational aspects of DEA. The researchers developed various methods and procedures that we can classify into three groups: preprocessing, linear programming methods, and big data methods. The computational experiments, for example [17][19][20], have shown that the savings achieved by applying some procedures presented in this paper can be significant.

Acknowledgements

The study was supported by the fund for long-term development of science and research at the Faculty of Informatics and Statistics Business (IP400040).

References

[1] Ali, A. I. (1993). Streamlined computation for data envelopment analysis. *European Journal of Operational Research*, *64*(1), 61–67. https://doi.org/10.1016/0377-2217(93)90008-b.

[2] Ali, A. I. (1994). Computational aspects of DEA. In Charnes, A., Cooper, W. W., Lewin, A. Y., Seiford, L. M., *Data Envelopment Analysis: Theory, Methodology, and Applications*, 63–88. https://doi.org/10.1007/978-94-011-0637-5_4.

[3] Andersen, P. & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, *39*(10), 1261–1264. https://doi.org/10.1287/mnsc.39.10.1261

[4] 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. https://doi.org/10.1287/mnsc.30.9.1078.

[5] Barr, R. S. & Durchholz, M. L. (1997). Parallel and hierarchical decomposition approaches for solving large-scale data envelopment analysis models. *Annals of Operations Research*, *73*:339–372.

[6] Charles, V., Aparacio, J. & Zhu, J. (2020). *Data Science and Productivity Analytics*. Springer.

[7] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

[8] Charnes, A., Cooper, W. W., Lewin, A. Y. & Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston: Kluwer Academic Publishers.

[9] Chen, Y. & Ali, A. I. (2002). Output–input ratio analysis and DEA Frontier. *European Journal of Operational Research*, *142*(3), 476–479. https://doi.org/10.1016/s0377-2217(01)00318-6

[10] Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). *Handbook on Data Envelopment Analysis*. Kluwer Academic.

[11] Deprins, D., Simar, L. & Tulkens, H. (1984). Measuring labor efficiency in post offices. In M. Marchand, P. Petieau, P. & H. Tulkens (Eds.), *The performance of public enterprises: concepts and measurement*. Amsterdam: North Holland.

[12] Dellnitz, A. (2022). Big Data Efficiency Analysis: Improved algorithms for data envelopment analysis involving large datasets. *Computers & Operations Research*, *137*, 105553. https://doi.org/10.1016/j.cor.2021.105553.

[13] Dlouhý, M., Jablonský, J. & Zýková, P. (2018). *Analýza obalu dat*. Professional Publishing.

[14] Dulá, J. H. (2002). Computations in DEA. *Pesquisa Operacional*, *22*(2), 165–182. https://doi.org/10.1590/s0101-74382002000200005.

[15] Dulá, J. H. (2011). An algorithm for data envelopment analysis. *INFORMS Journal on Computing*, *23*(2), 284–296. https://doi.org/10.1287/ijoc.1100.0400.

[16] Dulá, J. H., Helgason, R. V. & Hickman B. L. (1992). Preprocessing schemes and a solution method for the convex hull problem in multidimensional space. In O. Balci, O. (Ed.). *Computer science and operations research: new developments in their interfaces*. Pergamon Press; 59–70.

[17] Dulá, J. H. & López, F. J. (2009). Preprocessing DEA. *Computers & Operations Research*, *36*(4), 1204–1220. https://doi.org/10.1016/j.cor.2008.01.004.

[18] Emrouznejad, A. & Yang, G. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8. https://doi.org/10.1016/j.seps.2017.01.008.

[19] Khezrimotlagh, D., Zhu, J., Cook, W. D. & Toloo, M. (2019). Data Envelopment Analysis and big data. *European Journal of Operational Research*, *274*(3), 1047–1054. https://doi.org/10.1016/j.ejor.2018.10.044.
[20] Khezrimotlagh, D. & Zhu, J. (2020). Data Envelopment Analysis and Big Data: Revisit with a Faster Method. In V. Charles, J. Aparicio & J. Zhu (Eds.), *Data Science and Productivity Analytics*. International Series in Operations Research & Management Science, vol 290. Springer. https://doi.org/10.1007/978-3-030-43384-0_1.
[21] Sueyoshi, T. & Chang, Y. (1989). Efficient algorithm for additive and multiplicative models in data envelopment analysis. *Operations Research Letters*, *8*, 205-213.

The Crossing Numbers of Cartesian Product of Paths and Cycles with Several 8-vertex Graphs

Emília Draženská [1]

Abstract. The crossing number of a graph G is the minimum number of edge crossings over all drawings of G in the plane. Garey and Johnson have proved that compute the crossing number for a given graph is a very difficult problem, it is NP–complete problem, in general. There are several classes of graphs for which crossing numbers have been studied. The main aim of the paper is to establish the crossing numbers of the Cartesian products of the paths and cycles on n vertices with several connected graphs on eight vertices.

Keywords: graphs, drawings, crossing numbers

JEL Classification: C02
AMS Classification: 05C10, 05C38

1 Introduction

Let us consider a simple *graph* G whose vertex set is V and edge set E. A *drawing* D of the graph G is a representation of a graph G in a plane in such a way, that each vertex is represented by a point in $\mathbb{R}^2$ and each edge by a curve between its two endpoints. A *crossing* of two edges is the intersection of the interiors of the corresponding curves. The *crossing number*, $cr(G)$, of a graph G is the minimum number of edge crossings in any drawing of G in the plane. The drawing with a minimum number of crossings, must be a *good drawing*, that means, that in the drawing: (a) no edge crosses itself, (b) adjacent edges do not cross, and (c) no two edges cross more than once.

It is well known that the problem of reducing the number of crossings on the edges in the drawings of graphs was study in several areas, and the most significant area is very-large-scale-integration (VLSI) technology. Introduction of the VLSI technology revolutionized circuit design and had a strong influence on concurrent computing. The great deal of research aiming at efficient use of the newly discovered technologies has been done and further investigations are in advance. As a crossing of two edges of the communication graph asks unit area in its VLSI-layout, the crossing number and the number of vertices of the graph give a lower bound for the area of the VLSI-layout of the communication graph. The crossing numbers have been also studied to improve the readability of hierarchical structures and automated graph drawings. The visualized graph should be easy to read and understand. For the transparency of graph drawing, the reducing of crossings is relevant.

The investigation on the crossing number of graphs is a classical but very difficult problem. According to their special structure. Cartesian product of two graphs is one of classes of graphs for which the values of crossing numbers is studied. The *Cartesian product* $G_1 \Box G_2$ of two simple graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is a graph with vertex set $V = V_1 \times V_2$ and edge set $E = \{\{(x_1, y_1), (x_2, y_2)\}; x_1 = x_2$ and $\{y_1, y_2\} \in E_2$ or $y_1 = y_2$ and $\{x_1, x_2\} \in E_1\}$. That means, two vertices $(x_1, y_1), (x_2, y_2)$ are adjacent in $G_1 \Box G_2$ if their first coordinates are the same and their second coordinates are adjacent in G_2, or if their second coordinates are the same and their first coordinates are adjacent in G_1.

Let C_n be the cycle contains n vertices, P_n be the paths contains $n + 1$ vertices and S_n be the star isomorphic to complete bipartite graph $K_{1,n}$. The researchers are trying to determine the value of the crossing numbers for the Cartesian products $G \Box C_n$, $G \Box P_n$ and $G \Box S_n$ for some specific graphs G. Beineke and Ringeisen in [1], Jendrol' and Ščerbová in [8], Klešč in [9], Klešč and Kocúrová in [11], Klešč and Draženská in [4], Klešč and Petrillová in [12], Draženská in [2], [3] determined the crossing numbers of the Cartesian products of several graphs on at most eight vertices with cycles, paths and stars. Harary et al. [7] conjectured that the crossing number of $C_m \Box C_n$ is $(m - 2)n$ for all m, n satisfying $3 \leq m \leq n$. It was proved that the crossing number of $C_m \Box C_n$ equals its long-conjectured value at least for $n \geq m(m + 1), m \leq 7$ (see [6]).

In this paper we give the exact values of the crossing numbers for Cartesian products of cycles or paths with several hraphs of order eight. We consider graphs G_i, $i = 1, 2, \ldots 20$, on eight vertices and with eight edges containing a cycle C_3 or a cycle C_4, which are collected in Figure 1.

[1] Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, emilia.drazenska@tuke.sk

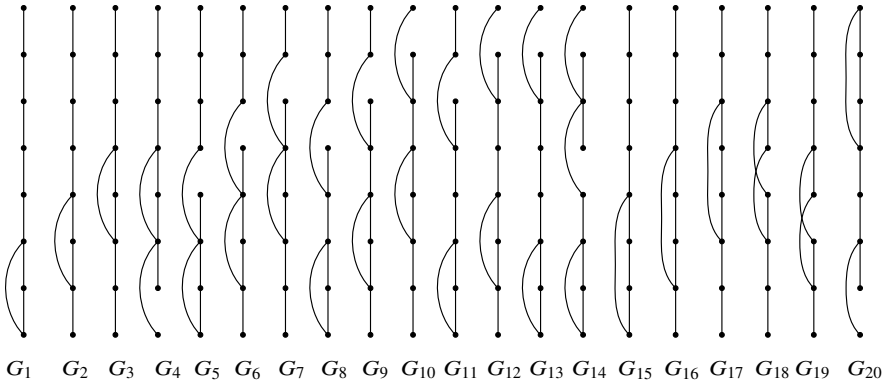


Figure 1 Several 8–vertex graphs with 8 edges with 3–cycle or 4–cycle

2 The Crossing Number of the Cartesian Products of some Graphs with Cycles

Every graph $G_i \square C_n$ contains n copies of the graph G_i. An uppper bound for the crossing number of the graph $G_i \square C_n$ is a number of crossings in a drawing of $G_i \square C_n$. On the other hand, if the graph $G_i \square C_n$ contains a subgraph for which the crossing number is known yet, we have a lower bound for the crossing number for the graph $G_i \square C_n$.

Theorem 1. *For $n \geq 6$,*

$$\begin{aligned}
&\mathrm{cr}(G_1 \square C_n) = \mathrm{cr}(G_2 \square C_n) = \mathrm{cr}(G_3 \square P_n) = n,\\
&\mathrm{cr}(G_4 \square C_n) = \mathrm{cr}(G_5 \square C_n) = \mathrm{cr}(G_6 \square C_n) = \mathrm{cr}(G_7 \square C_n) = \mathrm{cr}(G_8 \square C_n) = \mathrm{cr}(G_9 \square C_n) = \mathrm{cr}(G_{10} \square C_n) =\\
&= \mathrm{cr}(G_{11} \square C_n) = \mathrm{cr}(G_{12} \square C_n) = \mathrm{cr}(G_{13} \square C_n) = \mathrm{cr}(G_{15} \square C_n) = \mathrm{cr}(G_{16} \square C_n) = \mathrm{cr}(G_{17} \square C_n) =\\
&= \mathrm{cr}(G_{18} \square C_n) = \mathrm{cr}(G_{19} \square C_n) = 2n,\\
&\mathrm{cr}(G_{14} \square C_n) = \mathrm{cr}(G_{20} \square C_n) = 3n.
\end{aligned}$$

Proof: In Figures 2(a), 2(b) and 2(c) there are the drawings of the graphs $G_1 \square C_n$, $G_2 \square C_n$ and $G_3 \square C_n$ with n crossings. Thus, $\mathrm{cr}(G_1 \square C_n)$, $\mathrm{cr}(G_2 \square C_n)$ and $\mathrm{cr}(G_3 \square C_n)$ are at most n. The graphs G_1, G_2 and G_3 contain the star S_3 as a subgraph. So, all Cartesian products $G_1 \square C_n$, $G_2 \square C_n$, $G_3 \square C_n$ contain $S_3 \square C_n$ as a subgraph. Since $\mathrm{cr}(S_3 \square C_n) = n$ for $n \geq 6$ (see [1]), the crossing numbers of $G_1 \square C_n$, $G_2 \square C_n$ and $G_3 \square C_n$ are at least n for $n \geq 6$. Hence, $\mathrm{cr}(G_i \square C_n) = n$ for $i = 1, 2, 3$.

In Figures 2(d)–2(m) and 2(o)–2(s) are drawings of the graphs $G_i \square C_n$ for $i = 4, \ldots, 13$ and for $i = 15, \ldots, 19$ with $2n$ crossings and therefore, $cr(G_i \square C_n) \leq 2n$ for $i = 1, \ldots, 13, 15, \ldots, 19$. The graphs $G_i \square C_n$ for $i = 4, 5, 6, 7$ contain the graph $S_4 \square C_n$ as a subgraph. As $\mathrm{cr}(S_4 \square C_n) = 2n$ for $n \geq 6$ (see [1]), $\mathrm{cr}(G_i \square C_n) \geq 2n$ for $i = 4, 5, 6, 7$. The graphs G_i for $i = 8, 9, 10$ contain the graph T (see Figure 3) as a subgraph and the graphs G_i for $i = 11, 12, 13$ contain the subdivision of the graph T as a subgraph. In [4] was determined that $\mathrm{cr}(T \square C_n) = 2n$ for $n \geq 6$. Hence $\mathrm{cr}(G_i \square C_n) \geq 2n$ for $i = 4, \ldots, 13$.

The drawings of the graphs $G_i \square C_n$ for $i = 15, 16, 17, 18, 19$ with $3n$ crossings are shown in Figures 2(o)–2(s). Thus $\mathrm{cr}(G_i \square C_n) \leq 2n$ for $i = 15 \ldots, 19$. As $C_4 \square C_n$ contains all graphs $G_i \square C_n$, $i = 15 \ldots, 19$ and $\mathrm{cr}(C_4 \square C_n) = 2n$ for $n \geq 6$ (see [1]), the value $2n$ is the upper bound for crossing numbers of these graphs. Hence $\mathrm{cr}(G_{15} \square C_n) = \mathrm{cr}(G_{16} \square C_n) = \mathrm{cr}(G_{17} \square C_n) = \mathrm{cr}(G_{18} \square C_n) = \mathrm{cr}(G_{19} \square C_n) = 2n$.

As the graph G_{14} contains two vertex–disjoint subgraphs, one is isomorphic to C_3 and the second one, is isomorphic to S_4, we obtain $\mathrm{cr}(G_{14} \square C_n) \geq \mathrm{cr}(C_3 \square C_n) + \mathrm{cr}(S_4 \square C_n) = 3n$. And, the graph G_{20} contains C_4 and S_3 as a vertex–disjoint subgraphs, we have $\mathrm{cr}(G_{20} \square C_n) \geq \mathrm{cr}(C_4 \square C_n) + \mathrm{cr}(S_3 \square C_n) = 3n$. This confirms that $\mathrm{cr}(G_{14} \square C_n) = \mathrm{cr}(G_{20} \square C_n) = 3n$. This fact completes the proof. □

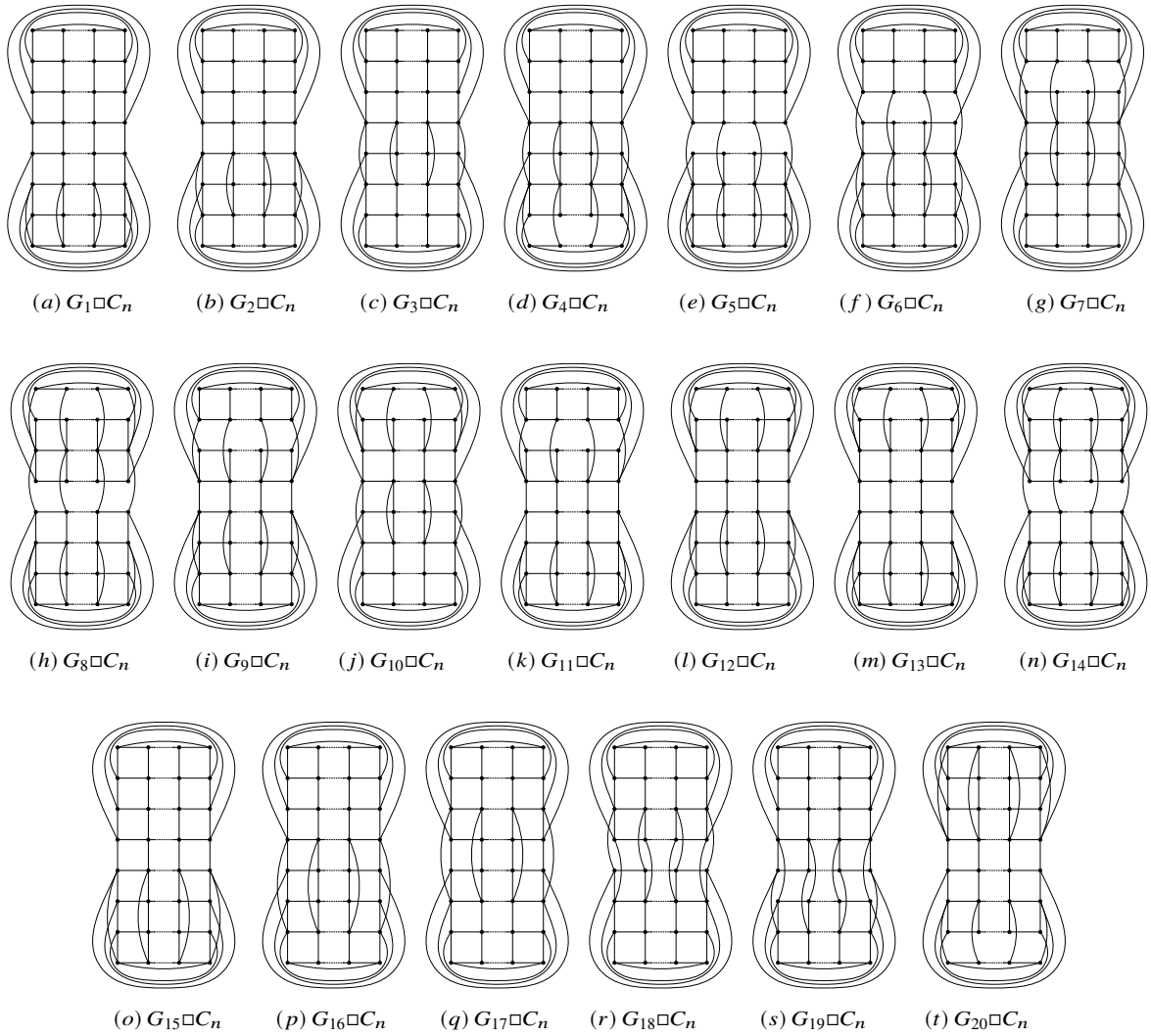


Figure 2 Cartesian products C_n with graphs G_i for $i = 1, 2, \ldots, 20$

Figure 3 The graph T

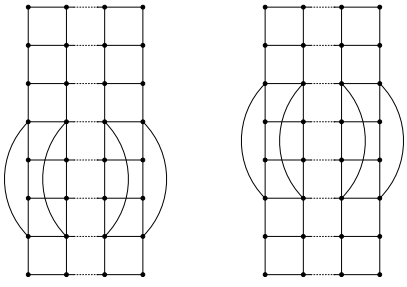


Figure 4 Cartesian products P_n with graphs G_i for $i = 16, 17$

3 The Crossing Number of the Cartesian Products of some Graphs with Paths

Every graph $G_i \square P_n$ contains $n+1$ copies of the graph G_i. Again, we will use the diagram of $G_i \square P_n$ to determine the upper bound for the crossing number of the graph $G_i \square P_n$ and the already published result to determine the lower bound for the crossing number of the graph $G_i \square P_n$.

Theorem 2. *For $n \geq 1$,*

$$\mathrm{cr}(G_{16} \square P_n) = \mathrm{cr}(G_{17} \square P_n) = 2(n-1).$$

Proof: In Figures 4(a) and 4(b) one can find the drawings of the graphs $G_2 \square P_n$ and $G_3 \square P_n$. The edges of $n-1$ copies of the graphs G_{16} and G_{17} are crossed twice, so the crossing number of $G_2 \square P_n$ and $G_3 \square P_n$ is at most $2(n-1)$.

Now, we find the lower bounds of crossing numbers $\mathrm{cr}(G_{16} \square P_n)$ and $\mathrm{cr}(G_{17} \square P_n)$. The graphs G_{16} and G_{17} contain the graph T as a subgraph. So, the Cartesian products $G_2 \square P_n$ and $G_3 \square P_n$ contain $T \square P_n$ as a subgraph. It was proved in [12] that $\mathrm{cr}(T \square P_n) = 2(n-1)$. It implies, that the crossing number of $G_2 \square P_n$ and $G_3 \square P_n$ is at least $2(n-1)$.

The upper and lower bounds of crossing numbers for graphs $G_i \square P_n$ for $i = 2, 3, 6, 8$ are the same. Thus, we get exact values of crossing numbers of corresponding graphs. □

References

[1] Beineke L. W. & Ringeisen, R. D. (1980). On the crossing numbers of products of cycles and graphs of order four. *J. Graph Theory*, 4, 145–155.

[2] Draženská, E. (2014). The crossing numbers of products of paths with 7-vertex trees. *Creative Mathematics and Informatics*, 23, 109–119.

[3] Draženská, E. (2014). The crossing numbers of several graphs of order eight with paths. *Mathematical modelling and geometry*, 1, 13–24.

[4] Draženská, E. & Klešč, M. (2007). The crossing numbers of products of cycles with 6–vertex trees. *Tatra Mountains Math. Publ.*, 36, 109–119.

[5] Garey, M. R. & Johnson, D. S. (1983). Crossing number is NP-complete. *SIAM J. Algebraic and Discrete Methods*, 4, 312–316.

[6] Glebsky, L. Y. & Salazar, G. (2004). The crossing number of $C_m \times C_n$ is as conjectured for $n \geq m(m+1)$. *J. Graph Theory*, 47, 53–72.

[7] Harary, F., Kainen, P. C. & Schwenk, A. J. (1973). Toroidal graphs with arbitrarily high crossing numbers. *Nanta Math.*, 6, 58–67.

[8] Jendroľ, S. & Ščerbová,M. (1982). On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$. *Časopis pro pěstování matematiky*, 107, 225–230.

[9] Klešč, M. (1991). On the crossing numbers of Cartesian products of stars and paths or cycles. *Mathematica Slovaca*, 41, 113–120.

[10] Klešč, M. (2005). Some crossing numbers of products of cycles. *Discussiones Mathematicae - Graph theory*, 25, 197–210.

[11] Klešč, M. & Kocúrová, A. (2007). The crossing number of products of 5-vertex graphs with cycles. *Discrete Mathematics*, 307, 1395–1403.

[12] Klešč, M. & Petrillová, J. (2013). The crossing numbers of products of paths with graphs of order six. *Discussiones Mathematicae - Graph theory*, 33, 571–582.

Traffic Flow Control Using Tecnomatix Plant Simulation

Jan Fábry[1], Ondřej Kopčan[2]

Abstract. The traffic jams that form as a result of poor control of traffic flows cause delays in the delivery of materials and goods, delays of public transport, but also a high production of emissions and noise. The paper focuses on the issue of traffic flow control at intersections. Computer simulation is a suitable tool for performing the analysis of traffic flow problems to find the efficient solution. In many cases it is necessary to decide on the type of intersection for certain location. As the alternatives, non-signalized or signalized crossroad can be considered, or roundabout can be built. We created several simulation models in Tecnomatix Plant Simulation software to run computer experiments. In case of signalized intersections, experiments for intervals setting were executed.

Keywords: traffic flow, non-signalized intersection, signalized intersection, traffic lights, roundabout, computer simulation, Tecnomatix Plant Simulation software

JEL Classification: C63
AMS Classification: 90B20

1 Introduction

Due to the ever-increasing traffic growth in cities, there is a need to address the current problems of traffic flow control. Even in the age of modern digital technology and advanced communication systems, it is necessary to use established methods based on mathematical modelling and computer simulation that is the appropriate tool for adjustment of complex dynamic and stochastic systems.

Traffic flow problem can be considered from both transportation planning and traffic engineering perspectives [5]. The paper focuses on the intersections to improve traffic flows in the time period of traffic jams in cities. Three basic structures of street intersections can be considered: non-signalized intersections, signalized intersections and roundabouts. There are the analytical optimization approaches to control flows using traffic lights. In [9] the *genetic algorithm* is proposed to minimize the length of queues in front of traffic lights. Optimal traffic flow control on *roundabouts* is discussed in [8]. The authors consider two cost functionals: the total travel time and the total waiting time, which give an estimate of the time spent by drivers on the network section. In the survey, the cost functionals are computed analytically for a single intersection and the control parameters that locally optimize the flow are given. Traffic flow control is optimized also in [5].

The use of the *neural networks* for the analysis traffic flows is demonstrated in [10]. Authors use the neural network to predict the changes of traffic flow in the future. Predicted data is compared with the actually observed traffic flows. The research had great significance to the future intelligent traffic decision-making.

The paper [2] deals with the description of the simulation transport hub model in the urban concentration of the city in Slovakia. The authors developed the *simulation model* to analyze the transport hub, which is the hot traffic bottleneck in the city. The simulation model created in EXTENDSIM software is used to verify the individual variants. The development of a simulation-based dynamic traffic assignment procedure for mixed traffic flow conditions is reported in [3]. Several different physical vehicle types are explicitly considered and modeled, including cars, buses, motorcycles, and trucks. In addition, different behavioral rules, pre-specified-path driver, user-equilibrium driver, system-optimization driver, and real-time information driver, are considered in the solution procedure.

A *simulation multi-agent model* of traffic flows has been developed in [4]. An algorithm was developed for designing a software product that simulates a transport system, taking into account the movement of cars along the lanes and their behavior at the intersection with *traffic lights regulation* (*signalized intersections*). The study [6] attempts to analyze traffic delay and queue length at roundabouts in Japan by computer simulation. The case study of traffic flow performed in Malaysia uses computer simulation software AASIDRA.

[1] Prague University of Economics and Business, W. Churchill Sq. 1938/4, Prague, fabry@vse.cz
[2] Logio, Evropská 2588, Prague, kopcan@logio.cz

In the paper, three types of intersections are considered: a non-signalized intersection, a signalized intersection and a roundabout. In case of non-signalized and signalized intersections, we focused on T-shaped intersections. Pedestrians and bikers are ignored in the analysis. For this purpose, many simulation programs specialized for traffic flow can be used, e.g. VISSIM, SUMO, AIMSUN, TRANSMODELER, PARAMICS, CITYFLOW etc. The aim of the survey is to make computer experiments with the simulation models created in the TECNOMATIX PLANT SIMULATION software developed by Siemens [1]. This discrete-event simulation system is primarily used for modelling, analyzing, and optimizing complex production systems and logistics processes. It is commonly used in manufacturing industries to design and optimize production systems, material handling systems, and supply chain networks. However, as the next text shows, the complexity and flexibility of the TECNOMATIX PLANT SIMULATION environment and the advanced simulation possibilities allow the computer simulation of traffic flows to a relatively high level of detail even in this software.

2 Tecnomatix Plant Simulation Software

As mentioned previously, the Tecnomatix Plant Simulation (TPS) [1] is primarily used for modelling and optimizing production and manufacturing systems, e.g. batch control or the actual planning of the production program structure. In the paper, we show the possibilities of using the software for traffic flow simulation. Discrete-event simulation of intersections flow with the object-oriented programming language SimTalk [1] offers a versatile application, encourages a systemic approach to problem solving, and to a large extent eliminates the risks of implementing changes.

All models in the TPS 16.1 environment use basic mobile units (MUs). These have clearly defined attributes, a point of origin and a point of exit (*Source* and *Drain*). These two types of objects define system boundaries of the simulation model. The movement of MUs represents the flow from one object to another according to predefined rules. The TPS distinguishes three types of MUs: *part*, *container* and *transporter*. In our models, due to the specific focus on traffic flows, we use MUs of the *transporter* type. These have specific behavioral attributes determining their passing through the model (e.g., *speed*, *acceleration*, *length* or *direction of movement*). MUs of type *transporter* move on passive object of type *track*, which represents traffic path or road. By default, *tracks* have a capacity set to infinity and the maximum number of MUs that can move along it is given as the ratio of the length of the *track* object and the length of the *transporter*. *Track* may also contain *sensor*s, either relatively or absolutely distant from the start of the object. A *sensor* represents a decision point in the model that can be handled by a *method* object representing a function or a procedure to local or partial control of the simulation run. In reality, it could represent, e.g., inductive loop for traffic detection in front of the intersection.

For the signalized intersections, *generator* is the main control object. It sends impulses to the control method at a user-defined time intervals. The method is used for the regular set-up of traffic lights in all directions. *Experiment manager* is an object providing the user with the ability to perform experiments on the model created. Since the aim of the paper is to verify the general possibility of using the TPS for the traffic flow, we developed all models for the hypothetical intersections instead of real traffic systems. Therefore, all the simulation models were created for the methodological purposes.

3 Simulation Models of Intersections

A common basis, i.e., invariable specifics, is established for the possible comparison of all models. All the models previously mentioned represent a three-direction intersection, they are treated as discontinuous (isolated) solutions with no overlap with the other intersections in the traffic network. The models deliberately abstract from externalities affecting intersection flow in the real world, e.g., pedestrian control, different speeds (different types of vehicles), congestion peaks, vehicle accidents, etc. The generation of MUs (vehicles) is kept identical for comparability of results for all models and represents the arrival of a vehicle from another node in the traffic network. Similarly, the exit of vehicles represents leaving the intersection. Between the *source* and the *drain*, which form the system boundaries of all models, *track* objects are used to represent the roads. In addition, models contain decision points (e.g., traffic lights) and *methods* that form the core of each model and control the movement of vehicles. The simulation time is set the same for all models at one day. It should be emphasized that the simulation time does not reflect the changes of traffic flow in real time, i.e., different arrival rates in all directions through daytime. One day was set because of statistical significance of the results.

The vehicles generation itself occurs in intervals that vary for each direction to get the model closer to the real situations. For the purpose of correctness of the models, the strategy setting on the exit is maintained as blocking, which guarantees the preservation of vehicles that do not manage to pass through the model, e.g., due to insufficient

capacity of the road or intersection, traffic light phases, etc. This prevents the comparison results from being biased by vehicles that do not pass through the entire model or do not reach the intersection at all.

In order to be able to compare results throughout all models, the models observe two main variables:

1. The number of vehicles passing through the intersection (intersection volume).
2. Average waiting time.

The models use a master table to monitor all the vehicles (that want to pass through the intersection) and a unique ID for each vehicle. Using methods, the ID is supplemented with other key attributes in different parts of the simulation (time and place of creation, waiting time, average time spent in the model, time and point of exit). After finishing the simulation run, the data in the master table is used for the statistical evaluations.

The simulation of intersection passage in our models is determined by the flexible vehicles generation in all models, a well-defined interface for each model and a specific *method* that controls the passage of vehicles through the model (passage of vehicles through a given type of intersection). The following intersection types have been selected, especially in terms of commonly used approaches in the regions:

1. non-signalized T-shaped intersection;
2. roundabout;
3. signalized T-shaped intersection.

Non-signalized T-shaped intersection

The traffic flow control at individual intersections reflects reality as much as possible. The non-signalized T-shaped intersection (see Figure 1) is controlled as a major and a minor road, whereby vehicles from the minor have to give way to vehicles on the major road, just as left turning vehicles give way to the straight direction. All feasible vehicle movements are controlled with the use of *methods*.

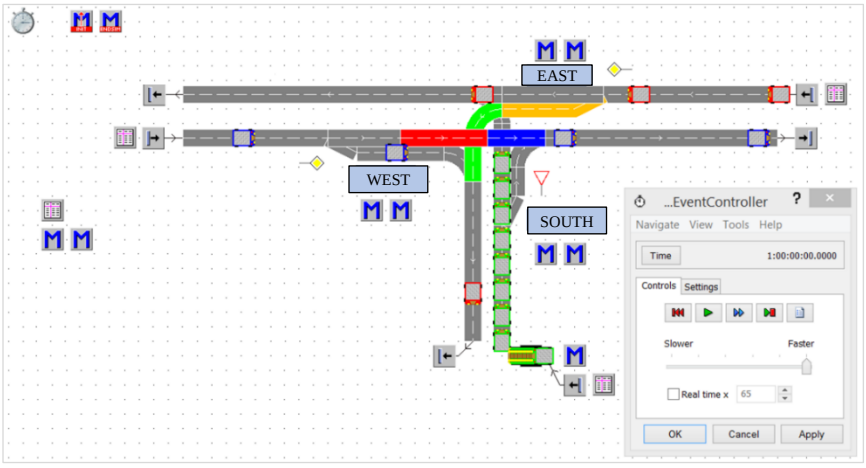


Figure 1 Simulation model of the non-signalized T-shaped intersection

Roundabout

The roundabout (see Figure 2) is modelled according to regional practice, meaning vehicles entering the roundabout give way to vehicles travelling on the roundabout. In the simulation model, *methods* are again used for the control of feasible vehicle movements. For this type of intersection, we have investigated two designs, namely a small and a large roundabout. The size of the roundabout plays a key role in this type of intersection, thus affecting the time vehicles spend in the intersection and the capacity of the intersection itself. A larger roundabout also requires a larger parcel of land and a larger financial investment associated with the need for more complex civil engineering modifications.

Signalized T-shaped intersection

For an intersection controlled by a traffic lights (see Figure 3), the *method* that controls the flow logic is crucial. In our model, the main control element is the *generator* object, which sends impulses to the control *method* at a user-defined time. The impulses interval is set to 1 second.

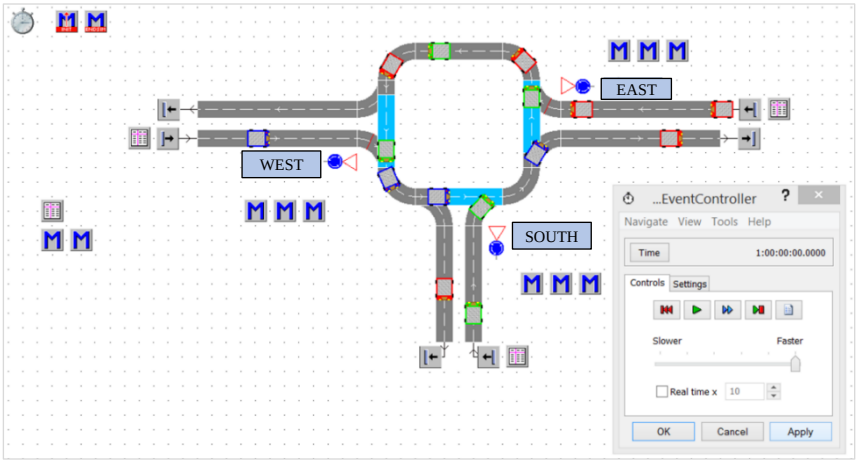


Figure 2 Simulation model of the (small) roundabout

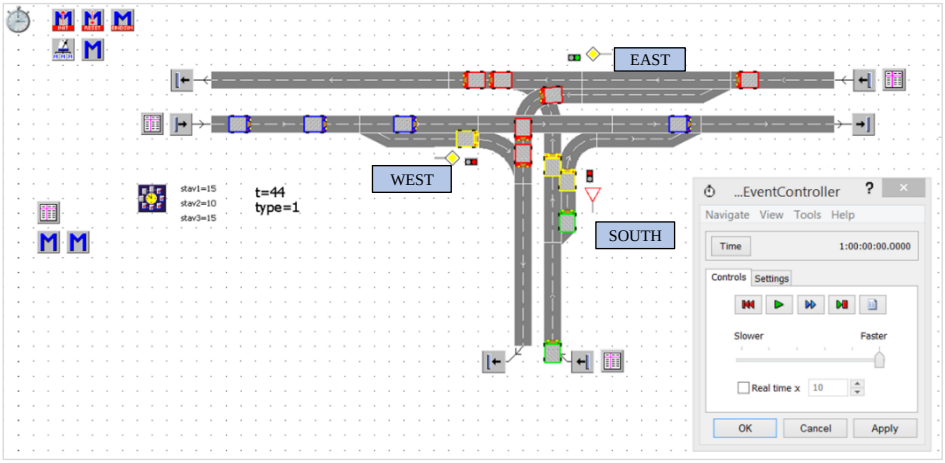


Figure 3 Simulation model of the signalized intersection

Methods

Because the *method* for traffic lights control is the core of the model, we offer the main part of it:

```
// if cycle for the variable t, which is used to set the total time for controlling the intersection
// the sum of the signal go and clearing times (4s*3) is assigned to the variable time
time:=state1+state2+state3+(3*4);
if t=0 then
    t:=time;
else
    t:=t-1;
end;

// if cycles for setting time intervals for individual signal groups
// the output of this part of the program is key for the creation of the signal program
if t>time-stav1 AND t≤time then
    prom:=time;
end;
```

```
if t>time-stav1-4 AND t≤time-stav1 then
    prom:=time-stav1;
end;
if t>time-stav1-4-stav2 AND t≤time-stav1-4 then
    prom:=time-stav1-4;
end;

// if cycle that examines the current value of the variable t and sets one of the 4 signal phases
// at time t, which is equal to one of the control methods, the whole command is executed
// the program then waits for re-execution or for the next value of variable t
if prom=time then
    type:=1; // signal phase #1
end;

// NORTHBOUND STOP
Track12.EntranceLocked:=true;
Track15.EntranceLocked:=true;
TrafficLight_N.CurrIconNo:=3; //red light

// WESTBOUND GO
Track32.EntranceLocked:=false;
Track36.EntranceLocked:=false;
TrafficLight_W.CurrIconNo:=2; //green light

// EASTBOUND STOP
Track22.EntranceLocked:=true;
Track26.EntranceLocked:=true;
TrafficLight_E.CurrIconNo:=3; //red light
```

Experiments

The intersection control uses a total of 4 phases - 1 for each of 3 directions and 1 for clearing phase. The clearing phase is user-fixed at 4 seconds, the other phases are dynamic through 3 unique variables. By means of experiments in the TPS environment, it is possible to use dynamic variables to find the optimal variant for the intersection. The source for the experiment is a table of all possible combinations to build a signal program, i.e., a variation of the k-th class ($k = 3$) with repetition of n elements ($n = 4$) means 64 combinations (user-defined intervals of 10, 15, 20, 25 seconds for 3 directions). The output parameters are the main variables of the model (number of passes and the average waiting time). The significance level is set to 95 %, and a total of 4 observations are performed per each experiment, for a total of 256 experiments.

For the simplicity, in the presentation of results we will concentrate on the values of the average waiting time. To be more realistic, a method of generating vehicles on a negative exponential distribution with mean was chosen. We assume the intervals with means (West 9s – South 13s – East 10s). In Table 1 the results for non-signalized intersection are given (average waiting times are calculated in seconds).

Non-signalized intersection	West	South	East
West	-	24.6	5.2
South	55.4	-	36.1
East	3.3	27.1	-

Table 1 Average waiting times for non-signalized intersection

The high value of the average waiting time for vehicles going from the south to the west can be explained by the frequent necessity of their giving way to other vehicles. Low values for vehicles going from the west to the east and from the east to the west are also understandable. Non-zero values correspond to the fact that vehicles can be stuck because of preceding turning vehicles.

The roundabout improves the passage through the intersection in the sense of correcting the disbalance of the traffic flow caused by the major and minor roads (see Table 2). The different values of the average waiting time are mainly affected by different mean of interarrival times for directions. Large roundabouts give even better values because of driver's convenient visibility and due to easier and safer acceleration. On the other hand, in use of large roundabout, vehicles are naturally determined to spend more time in the intersection itself – due to more distance to be covered.

Small roundabout	West	South	East
West	-	27.2	26.3
South	13.2	-	14.1
East	21.1	20.3	-

Table 2 Average waiting times for small roundabout

If we restrict ourselves to the variable of the average waiting time, the use of a non-signalized T-shaped intersection or a small roundabout is inappropriate for the traffic volume we have determined. On the other hand, the large roundabout and signalized intersection proved to be suitable, with minimal differences in the observed variable. Experiments were performed for various means of interarrival times.

The signalized intersection is always investigated using experiments and set to the effective variant in terms of the main variable of interest. The individual phases are artificially extended by an interval of 5s and are intended to reflect the effects of turnouts, driver inattention, etc. Similarly, it reflects the imperfection of traffic lights control, which in the real world does not examine the last vehicle from a given direction, but simply follows a given signal plan.

Generation of vehicles in means of interarrival time (9s – 13s – 10s) and traffic lights settings (20s – 15s – 20s) lead to the outputs in Table 3. The chosen traffic volume is so low that the models accommodate all vehicles without much delay.

Signalized intersection (20s – 15s – 20s)	West	South	East
West	-	27.0	27.5
South	26.2	-	25.2
East	25.9	37.1	-

Table 3 Average waiting times for signalized intersection (20s –15s – 20s)

For comparison, Tables 4 and 5 show the results for generation of vehicles in means of interarrival time (9s - 13s - 10s) and traffic lights settings (15s – 10s – 12) and (10s – 10s – 10s), respectively. The results for the last settings are comparable with the results for a small roundabout. If the traffic flow increased, we could observe worse values in the model with a small (or large) roundabout that could no longer accommodate vehicles with its capacity, while traffic lights would enable the appropriate control of traffic flow.

Signalized intersection (15s – 10s – 12s)	West	South	East
West	-	25.1	25.5
South	21.9	-	20.7
East	21.6	21.7	-

Table 4 Average waiting times for signalized intersection (15s –10s – 12s)

Signalized intersection (10s – 10s – 10s)	West	South	East
West	-	25.0	23.5
South	18.2	-	17.2
East	21.1	22.0	-

Table 5 Average waiting times for signalized intersection (10s –10s – 10s)

4 Conclusion

Abstracting from other externalities and criteria, the dynamic models we built clearly identified two solutions that could be elaborated in more detail to find the optimal variant. The suitability of using a dynamic simulation tool to support decision making is clearly demonstrated here. In reality, we would expect other variables and externalities to be investigated and processed, e.g., the economic aspects of the project, the realistic possibilities of engineering modifications, the size of the used area or the criterion of traffic safety.

Modelling and using dynamic computer simulation, it is impossible to reflect reality 100%, and certain level of abstraction is always necessary. Thanks to the models, we have built, and the experiments, we have carried out, we have reached a partial conclusion. In our traffic flow models, the level of abstraction determines the final results. Our original hypothesis that a small roundabout would be more appropriate in optimizing traffic flows was not confirmed. This is because the model does not reflect a key factor, namely driver behavior. In real-world conditions, drivers may instinctively evaluate the appropriate moment to enter the intersection (e.g., due to better visibility, safer and less risky entry, slower vehicles on the road, etc.). In addition, real larger roundabouts have obviously two lanes which improve the traffic flow. Experiments show that roundabouts can be successfully used when the traffic flows are balanced in terms of all directions, while signalized intersections are particularly suitable in situations where traffic flows differ for each direction.

Additional objective of the paper was to show the flexibility of the Tecnomatix Plant Simulation software which is primarily designed for simulating inbound production processes. We have confirmed that the software can be used also outside the main focus. The traffic flow models we have created use objects and environmental elements in specific models and thus clearly demonstrate the complexity and general use of the software.

As to the future research we can consider the following improvements to presented simulation models:

- including pedestrians as additional MUs and adjusting their arrival rates;
- definition different speeds for MUs;
- considering other types of intersections (X-shaped intersections, 4+ directions roundabouts, etc.);
- incorporating the intersection into a comprehensive traffic network (holistic view of traffic flow);
- definition of a destination matrix to model traffic volumes from different directions;
- definition of the traffic volumes based on real data;
- consideration of daytime traffic flow fluctuations.

All those improvements should bring all models closer to real traffic flow systems.

Acknowledgements

Supported by the Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] Bangsow, S. (2020). *Tecnomatix plant simulation : modeling and programming by means of examples*. Second edition. Cham, Switzerland: Springer Nature, 816 pp.

[2] Fedorko, G., Rosova, A. & Molnar, V. (2014). The application of computer simulation in solving traffic problems in the urban traffic management in Slovakia. *Theoretical and Empirical Researches in Urban Management, 9*, 5-17.

[3] Hu, T. Y., Tong, C. C., Liao, T. Y. et al. (2018). Dynamic route choice behaviour and simulation-based dynamic traffic assignment model for mixed traffic flows. *KSCE Journal of Civil Engineering, 22*, 813–822.

[4] Kasatkina, E. V. & Vavilova, D. D. (2020). Computer simulation of traffic flows. *Journal of Physics: Conference Series*.

[5] Khan, S., Ahmed, A., Nasir, H. et al. (2018). Optimal traffic flow control on roundabouts: A review. *KSCE Journal of Civil Engineering, 22*, 752-762.

[6] Lu, W., Vandebona, U., Kiyota, M. et al. (2020). Estimation of Traffic Delay at an Unconventional Roundabout by Computer Simulation. *IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Dalian, China, 295-302.

[7] Mohammed, A. (2013). Traffic Flow Analysis for Intersection Using Computer Simulation AASIDRA Software: A Case Study in Bangi Malaysia. *Tikrit Journal of Engineering Sciences, 20*. 10-25.

[8] Obsu, L. L., Monache, M. L. D., Goatin, P. et al. (2015). Traffic Flow Optimization on Roundabouts. *Mathematical Methods in the Applied Sciences. 38*. 3075-3096.
[9] Teo, K. T. K., Kow, W. Y. & Chin, Y. K. (2010). Optimization of Traffic Flow within an Urban Traffic Light Intersection with Genetic Algorithm. *Second International Conference on Computational Intelligence, Modelling and Simulation*, Bali, Indonesia, 172-177.
[10] Wang, Z. & Zhao, L. (2019). Analysis and prediction of traffic flow based on Wavelet-BP neural network. *Journal of Physics: Conference Series, 1325*.

Model of a Network Industry

Petr Fiala[1]

Abstract. Behaviour in network industries is an important subject of intensive economic research. The analysed model of a network industry is based on the co-opetition concept, where players are networks that compete and networks consist of a supplier and producers who may or may not cooperate with the supplier. The effects that arise depending on cooperation and competition in networks are analysed. This simplified situation allows greater transparency of results. Cournot oligopoly model serves as a basis that is analysed as a game where strategic decisions are based on output. The other approach is based on negotiations between suppliers and non-cooperating producers. Supply prices are negotiated between the supplier and individual producers. The negotiation process can be modelled by a simultaneous biform game with a combination of cooperative and non-cooperative approaches. The model is simple, but it still allows to make important managerial implications.

Keywords: network industry, game theory, co-opetition

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Many modern industries involve networks (see e.g. [20], ([8]). In some cases, these networks are interconnected with others and engaged in competition for subscribers. The network industry typically requires access to rival networks to provide services or to satisfy its customers. Such examples include networks for communication services, electricity transmission, gas transportation, banks' ATM networks, etc. This feature is what distinguishes the network industries from others in that interconnected firms try to take dominant position not only by competing in prices but also by deteriorating competing network by charging excessive access fees. There are still many open research questions in the network competition analysis. Further development of network competition theory can help resolve these issues and improve policy (see [14]). The traditional network industries are in the areas of transport (see e.g. ([17]), (rail, road, air, maritime, urban), communications (mobile, postal services), (see e.g. ([12]), energy (electricity, gas), water and wastewater. Profound technological transformations, especially digitalization has furthermore transformed these traditional network industries and has given rise to a new type of network industries, namely digital platforms. These call for new and innovative ways of regulation. The motivation for this paper was to create a theoretical framework that would capture model tools for analyzing competition and regulation of both traditional and new network industries. Specific models analyze interconnections of network competition, regulation and integration.

2 Basic Theoretical Framework

The theoretical framework is based on traditional network competition literature e.g. ([1], [15], [16] and [7]] and is modified to meet the needs of specific models. The framework serves as a supply of methods and models from the areas of game, oligopoly, and negotiation theories. Laffont, Rey and Tirole (LRT) have analyzed a model of two local network companies that possess different attributes for consumers. In their model, the two companies, given access charges, set the local prices competitively. The customers of a network are charged the same price independent of the network which completes their call. The networks compete only in prices since the other attributes are assumed to be fixed. They use the standard Hotelling (1929) location model [13]. LRT model have provided a basic theoretical framework to analyze the network competition and interconnection issues.

The simplicity of the framework suggests that it should be possible to extend it in a number of different directions. The framework can be extended in terms of the number of networks, economic instruments, cost structures, price discrimination, asymmetric structures, etc. There is a vast literature about the possible extensions, e.g. ([4], [5]). The general framework contains the extension possibilities and modeling instruments.

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

The modeling instruments for analyses of network competition are:

- game theory models;
- oligopoly models;
- negotiation models.

Game theory is the basis for development of network competition models, non-cooperative and cooperative models as well. The Cournot, Stackelberg and Bertrand models are representations of oligopolistic behavior. Nash equilibrium concept is used for solution. Cartel models are representations of cooperative behavior. Negotiations take place in cooperative solutions of competition problems (e.g. see [19]). There are approaches based on game theory and other approaches including ones based on multicriteria evaluations. The modeling framework serves as a common basis for developing special models for analyzing specific features in network competition.

3 Co-opetition Modelling

The analysed model of a network industry is based on the co-opetition concept. In this section, the concept of co-opetition and information about its modelling using biform games is presented.

3.1 Co-opetition Concept

The co-opetition concept (see [2]) goes beyond the use of separate approaches to competition and cooperation by combining their advantages together. The corresponding model PARTS contains five items that form the basis of the functioning of the co-opetition concept in finding a solution to the analysed situation: Players, Added value, Rules, Tactics, Scope.

Players are divided by types into producers, customers, suppliers, competitors and complementors. The situation is analysed as a game between the relevant players. Expanding the number of player types from a wider environment provides a deeper analysis of the situation (relationships between suppliers and producers affect costs, relationships between producers and customers affect demand and prices, relationships between producers and competitors affect producer behaviour, relationships between producers and competitors bring mutual added value). The model should analyse the impulses that affect players to be competitors or to cooperate.

Producers receive resources from suppliers, they produce products that they supply to customers. Suppliers send resources to producers and receive money from them for those resources. Customers receive products from producers and send money back to producers. Producers compete to secure resources in quantity, price and quality. From the supplier's point of view, another producer is a competitor if it is more advantageous for the supplier to secure that competitor than the producer. From the customer's point of view, another producer is a competitor if it offers the substitution product to customers more advantageously. Complementors are the competitors whose products increase the value of the producer's product.

From the supplier's point of view, another producer is a complementor if it is more advantageous for the supplier to secure resources to the producer and also to other producers than if it is supplying the producer only. From the customer's point of view, another producer is a complementor if customers rate the producer product better if they also have products from other producers than if they only have the producer product. Complementors are inverse to competitors because higher demand for their products will lead to higher demand for the producer product. Players can play multiple roles simultaneously. A player can play the role of competitor and complementor at the same time.

Added values is provided by complementors if their products bring an extension of possibilities for producers' own products. The producers can recognize these added values and take activities that increase their profitability.

Rules create a structure of negotiation between players. Some rules are hard and cannot be changed during negotiations. Other rules are softer and can be changed in the negotiation of contracts.

Tactics are sequences of activities that form the monitoring of the negotiation process by other players. Players can use these certain activities to intentionally influence the behaviour of other players. It is useful to monitor these activities and respond to them accordingly.

Scope is determined by the interconnections between the PART elements of the game model and other possible games in which players from this model participate. Extending the scope with more games can increase profitability. Leaving games separate may prove advantageous if the interconnection would limit some businesses. Joining and separating games is determined by changes in conditions over time.

The current business is characterized by a rapid change in conditions. New products are coming to the market faster and faster. Discovering complementary relationships provides new opportunities to bring added values. This forms the basis of the concept of co-opetition. Therefore, it is necessary to include this dynamic in the new models of co-opetition. Cooperation models may include multiple criteria, such as economic, technological, social, environmental, and others, that evaluate new opportunities with complementors and the negotiation process with competitors. More sophisticated co-opetition models will be created by more consistent use of knowledge from economics, game theory, supply chain theory, and other disciplines. This paper attempts to use the co-opetition concept in modelling and analysis of network industries. The biform game models can be more adapted to respect the co-opetition concept with basic elements such as players, added values, rules, tactics, and scope. Players are represented by negotiators who use their tactics to look for added value. Players negotiate under the pressure of the rules with varying scope.

3.2 Biform Games

The co-opetition concept uses biform games (see [3]) as a combination of the non-cooperative and cooperative games theory. Modern game theory is based on the classic work of John von Neumann and Oskar Morgenstern [21]. Since then, many publications on game theory have been published. A good overview of models and solutions of non-cooperative and cooperative games is provided by Myerson's book [18]. Biform games can be divided into sequential and simultaneous types (see [11]).

Sequential Biform Games

The game consists of two phases. In one phase, one type of game models (cooperative or non-cooperative) is used, and in the other phase, the additional type following the previous part is used in the sequence.

For example, the sequential biform game is applied in supply chains (see [9]). The first phase is non-cooperative and addresses maximization of profit from customers with stochastic price-dependent demand and then division of the profit between producers and retailers based on the creation of a specific buy-back contract negotiations. The contract is easy to use, the chain delivers profit as a coordinated unit, and this profit can be arbitrarily distributed by setting a single parameter. The second phase is cooperative and concerns two issues. The first issue is the creation of a coalition of producers with regard to the capacity of resources. The second issue is the fair distribution of the total profit of producers to individual producers using the Shapley concept.

The reverse procedure can be used also. The first phase will be cooperative and the second will be non-cooperative. In the first phase, the maximum output for the coalition of all players is calculated and certain parameters are determined cooperatively with regard to the interests of the participants. In the second phase, players compete on other parameters of the model and the solution of the situation is determined non-cooperatively, for example using the concept of Nash equilibrium.

Simultaneous Biform Games

The simultaneous biform has a single phase in which cooperative and non-cooperative approaches are used simultaneously to find a cosensual solution of a situation. However, finding this consensual solution can take place through multi-round negotiations. The situation is influenced by the composition of the coalition of cooperating players and the level of their cooperation. Relationships between players can be cooperative or non-cooperative at the same time. For example, relationships between producers and complementors are competitive because they provide competitive substitution products and at the same time cooperative because they add value by expanding the use of competitors' products. Various restrictions affect players who are under pressure and thus determine the level of cooperation. The level of cooperation can change over time and can be measured by multiple criteria.

4 Model and Analysis of Network Industry

The model of a network industry is based on the co-opetition concept, where players are networks that compete and networks consist of a supplier and producers who may or may not cooperate with the supplier. The effects that arise depending on cooperation and competition in networks are analysed. A model is proposed that is based on simplifying assumptions in order to focus on the analysed effects.

4.1 Model

The structure of the network industry consists of m networks $N_i = (S_i, n_i, r_i)$, $i = 1, 2, \ldots, m$. The structure of the whole network industry system is shown in Figure 1.

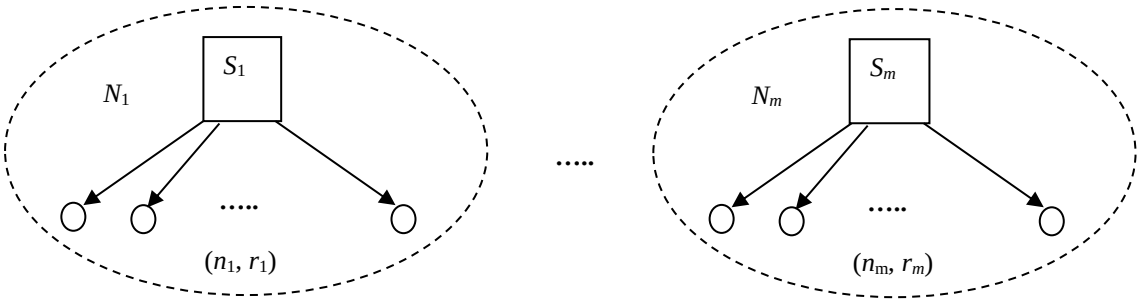


Figure 1 Structure of network industry

Each network N_i, $i = 1, 2, \ldots, m$, contains one supplier S_i, and n_i producers, of which r_i cooperates and $(n_i - r_i)$ does not cooperate with the supplier. There is a simple process of processing in networks. Suppliers sell one unit of input to producers, who process it into one final unit of output. Producers have limited capacity and can produce a single output. The outputs of producers are perfect substitutes.

Producers produce a total of n outputs, $n = \sum_{i=1}^{m} n_i$.

We assume a linear demand function and the price $p(n)$ of production is expressed as

$$p(n) = b - n, \tag{1}$$

where b is a parameter of the price function.

The following costs are assumed. For simplicity, variable costs are not considered. Each producer faces a fixed cost of size f. The supplier cooperates with some producers for strategic reasons. Each downstream unit has fixed costs. The cooperating producers have higher fixed costs with a fee of size g.

Then the profit for each non-cooperating producer is equal to $\pi = p - f$, and for each cooperating producer equals $\pi = p - f - g$. The profit that reaches the entire network N_i, $i = 1, 2, \ldots, m$, is equal to

$$\Pi_i(n_i) = p(n)\, n_i - f n_i - g r_i. \tag{2}$$

4.2 Analysis

The above network industry model can be analysed by various approaches. The first approach is based on the classical Cournot oligopoly model when the supplier has all the negotiation power. The concept of Nash equilibrium can be used for the non-cooperative model. Nash equilibrium is found by solving the following system of equations

$$\frac{\partial \Pi_i(n_i)}{\partial n_i} = 0,\ i = 1, 2, \ldots, m. \tag{3}$$

This system has a symmetric Nash equilibrium solution

$$n_i^* = \frac{1}{m+1}(b - f),\ i = 1, 2, \ldots, m. \tag{4}$$

This solution will be used for comparison with the solution obtained by negotiations.

The second approach is based on negotiations between suppliers and non-cooperating producers. Each supplier S_i, $i = 1, 2, \ldots, m$, will establish strong links with r_i cooperating producers and $(n_i - r_i)$ non-cooperating producers will enter the network to be secured by the supplier S_i. A supply contract is negotiated between the supplier and the non-cooperating producers.

Supply prices are negotiated between the supplier and individual producers. The supply prices $a_i(n_i, r_i)$ can also be interpreted as network access fees for producers. When negotiating, different producers may negotiate different prices, but due to the symmetry between them, in equilibrium, prices equalize. The negotiation process can be modelled by a simultaneous biform game with a combination of cooperative and non-cooperative approaches, the concept of pressures can be applied.

The decisions of the negotiating parties are influenced by internal and external pressures. The concept of pressure can be applied. The negotiations are about quantities n_i and transfer supply prices $a_i(n_i, r_i)$. The supply prices can

be interpreted also as access fees for downstream units to be members of the network. When negotiations between the supplier and a particular producer fall apart, the supplier must renegotiate price agreements with other producers and may face the competitive reaction of competing networks. A higher number of cooperating producers allows the supplier to negotiate higher access fees for non-cooperating producers and thus reduces the number of those who are willing to enter the network. The existence of cooperating producers reduces the negotiation power of non-cooperating units. The effect is so significant that the number of non-cooperating producers decreases more than the number of cooperating producers increases, and thus the total number of producers decreases.

All these pressures affect the negotiation power of the participants and thus change the trajectories of the effects of the pressure during the negotiation process. Only the existence of a possible cooperation changes the negotiation power of the supplier.

It can be shown (see [6]) that there is a symmetric Nash equilibrium solution for negotiation model without cooperating producers with the number of producers n_i^0, $i = 1, 2, \ldots, m$, for which it holds

$$n_i^0 = \frac{3}{3m+1}(b-f) > n_i^* = \frac{1}{m+1}(b-f), \quad i = 1, 2, \ldots, m. \tag{5}$$

We denote the resulting access fees for non-cooperating producers in case without cooperating ones as $a_i^0 = a_i(n_i, 0)$, $i = 1, 2, \ldots, m$. The question is how to set the access fee a_i^0 in comparison with fixed costs g.

The impetus for cooperation is given by the following relationships:

- if $g = 0$, there is total cooperation, $r_i = n_i = n_i^*$,
- if $g < a_i^0$, there is partial cooperation, $r_i < n_i$,
- for $g \geq a_i^0$, there is no cooperation, $r_i = 0$.

Cooperation will arise when the fee is low, and this will reduce network scope and overall performance. Cooperation will only be profitable if $g \leq \frac{1}{3m+1}(b-f)$. If the fee is higher, then no cooperation will arise.

It can be shown (see [6]) that there is a symmetric Nash equilibrium solution for negotiation model with cooperating producers with the number of producers n_i^C, $i = 1, 2, \ldots, m$, number of cooperating producers r_i^C, $i = 1, 2, \ldots, m$, for which it holds

$$n_i^C = min\left(\frac{b-f+2g}{m+1}, \frac{3(b-f)}{3m+1}\right), \tag{6}$$

$$r_i^C = max\left(0, \frac{b-f-(3m+1)g}{m+1}\right). \tag{7}$$

The model has important managerial implications. The competition and its impact on cooperation are analysed. Cooperation allows the supplier to negotiate higher access fees for non-cooperating producers. Higher competition between networks then reduces the likelihood of cooperation and can reduce the possibility of inefficient cooperation.

5 Conclusions

The paper presents a basic modelling framework for analyses of network co-opetition. The modelling framework is based on the concept of co-opetition which is modeled using biform games as a combination of non-cooperative and cooperative games. The biform games are divided into sequential and simultaneous according to the way of connecting non-cooperative and cooperative approaches of game theory. To seek consensus in simultaneous biform games, the use of pressure-based negotiation has been suggested.

The modelling framework provides the ability to create specific models for analyzing certain relationships in networks (see [10]). By connecting and disconnecting these specific models, it is possible to analyze a certain focused part of the interest within the network industry. Possible modifications of the model framework include the use of other general tools for modeling co-opetition on networks and procedures for their analysis, analyze various possibilities of interconnection of non-cooperative and non-cooperative parts, use multicriteria evaluation in general, other negotiation procedures, etc. Another important part is creating other specific models and considering ways to connect and disconnect specific models. Experiments with models can provide important managerial implications that can be translated into improvements in real policies affecting network industries.

Acknowledgements

This work was supported by the grant No. IGA F4/42/2021, Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] Armstrong, M. (1998). Network Interconnection in Telecommunications. *The Economic Journal, 108,* 545–564.

[2] Brandenburger, A. M. & Nalebuff, B. J. (2011). *Co-opetition.* New York: Crown Business.

[3] Brandenburger, A. & Stuart, H. (2007). Biform games. *Management science, 53,* 537-549.

[4] Chemla, G. (2003). Downstream competition, foreclosure and vertical integration. *Journal of Economics and Management Strategy, 12,* 261-289.

[5] De Fontenay, C. C. & Gans, J. S. (2004). Can vertical integration by a monopsonist harm consumer welfare? *International Journal of Industrial Organization, 22,* 821-834.

[6] De Fontenay, C. C. & Gans, J. S. (2005). Vertical integration and competition between networks. Review of Network Economics *4*(1): 4-18

[7] Dessein, W. (2003). Network Competition in Nonlinear Pricing. *RAND Journal of Economics, 34,* 593-611.

[8] Fiala, P. (2016). *Dynamic pricing and resource allocation in networks (in Czech)*. Praha: Professional Publishing.

[9] Fiala, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research, 24,* 267-281.

[10] Fiala, P. (2022). Modelling and analysis of co-opetition in network industries by biform games. *Central European Journal of Operations Research, 30,* 647–665.

[11] Fiala, P. & Majovská, R. (2020). Modeling the Design Phase of Sustainable Supply Chains. In S. A. R. Khan (Ed.), *Global Perspectives on Green Business Administration and Sustainable Supply Chain Management*. Hershey PA: IGI Global.

[12] Hotelling, H. (1929). Stability in Competition. *Economic Journal, 39,* 41-57.

[13] Houpis, G., Rodriguez, J. M., Serdarević G. & Ovington, T. (2016). The Impact of Network Competition in the Mobile Industry. *Competition and Regulation in Network Industries, 17,* 32-54.

[14] Knieps, G. (2016). *Network Economics: Principles - Strategies - Competition Policy*. Cham: Springer International Publishing.

[15] Laffont, J. J., Rey, P. & Tirole, J. (1998a). Network Competition: I. Overview and Nondiscriminatory Pricing. *RAND Journal of Economics, 29,* 1-37.

[16] Laffont, J. J., Rey, P. & Tirole, J. (1998b). Network Competition: II. Price Discrimination. *RAND Journal of Economics, 29,* 38-56.

[17] Laroche, F., Sys, Ch., Vanelslander, T. & Van de Voorde, E. (2017). Imperfect competition in a network industry: The case of the European rail freight market. *Transport Policy, 58,* 53-61.

[18] Myerson, R. B. (1997). *Game Theory: Analysis of Conflict.* Cambridge: Harvard University Press.

[19] Sauer, P., Dvorak, A. & Fiala, P. (1998). Negotiation between authority and polluters - Model for support of decision making in environmental policy. *Politická ekonomie, 46,* 772-787.

[20] Shy, O. (2001). *The Economics of Network Industries*. Cambridge: Cambridge University Press.

[21] von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Two-stage Efficiency Analysis Pitfalls

Lukáš Frýd , Ondřej Sokol

Abstract. DEA is one of the two main estimators of technical efficiency and is thus a popular tool in policy analysis of agricultural, environmental and other policies. Associated with these policies is the so-called two-stage efficiency analysis, where the first stage estimates DEA efficiency and the second stage estimates the effect of selected policy variables on DEA efficiency from the first stage. Although there is a large body of theoretical work dealing with the statistical properties of the DEA estimator, how to measure the inputs and outputs to the DEA model and the sensitivity in the second stage is completely neglected. We show that differently approximated albeit highly correlated basic inputs lead to specific technical efficiencies. This heterogeneity also leads to heterogeneity of results in the second stage of the analysis. Policy analysis built on a two-stage approach not only cannot be easily compared with each other, but may lead to erroneous decisions.

Keywords: data envelopment analysis, input measurement, efficiency analysis

JEL Classification: C50
AMS Classification: 90C90

1 Introduction

Efficiency estimation using the DEA method [2, 1], represents one possible approach in efficiency analysis. Due to the possibility to estimate efficiency for multiple inputs and especially for multiple outputs, DEA enables to analyze efficiency in the presence of negative outputs. This feature of the DEA estimator is currently widely used in efficiency analysis when one of the negative outputs is, for example, greenhouse gas. A central issue, especially for policy analysis, is the impact of policy measures on efficiency. For this purpose, a two-stage efficiency analysis is used. In the first stage, the technical efficiency of each unit is estimated. In the second stage, a regression analysis is performed to clarify which regressors have a positive or negative impact on efficiency. The results obtained are often causally interpreted. In this paper, we show that heterogeneity in the approximation of the fundamental inputs by strongly correlated covariates leads to different outputs in the second stage of the regression analysis. Specifically, we focus on the effect of heterogeneous measures of labour, capital, and additional inputs on the results of the impact of subsidies on the technical efficiency of farms in the Czech Republic, Poland, Slovakia, and Hungary in 2014-2019.

Specifically, the empirical literature often uses the amount of machinery, fixed capital, current assets, etc. as capital. Sokol and Frýd [11] show that although these inputs are strongly correlated, each of these inputs leads to a different efficient set and a different ordering of the estimated efficient units. This results in highly heterogeneous results from the second stage regression analysis. The current literature describes the statistical properties of the DEA estimator see [6, 10], or the behaviour of the DEA estimator under different specifications (see [3]). Efficiency is viewed as a single latent variable that represents technical efficiency. Due to the heterogeneity of the approximation of the basic inputs, we obtain different efficiency sets and hence different technical efficiencies. However, this problem is hardly addressed from a theoretical perspective and is completely neglected in the empirical literature.

2 Data and Methodology

2.1 Data

The analysis is based on an unbalanced panel of 995 farms from the Czech Republic, Poland, Hungary, and Slovakia over the period from 2014 TO 2019. The data we use come from *Farm Accountancy Data Network* (FADN). FADN is an agriculture database, maintained under auspices of the European Commission. FADN participation is voluntary for farms. Data from a sample of farms are sent to a national branch of FADN (so-called *liaison agent*), which transmit the data to the global FADN database and is responsible for the international comparability of the data, i.e., the methodology of the collecting the data shall be the same in every EU country.

Department of Econometrics, Prague University of Economics and Business, Winston Churchill Square 4, 13067 Prague, Czech Republic, lukas.fryd@vse.cz

Department of Econometrics, Prague University of Economics and Business, Winston Churchill Square 4, 13067 Prague, Czech Republic, ondrej.sokol@vse.cz

The survey does not cover all the farms in the Union but only those which due to their size could be considered commercial. In total, the FADN sample consists of about 80 000 holdings and represent about 5 million farms using about 90 percent of the total utilized agricultural area and producing about 90 percent of agricultural marketable output. The database consists of ~ 5000 economic and other variables on an annual basis.

2.2 Efficiency Estimation

We use DEA models with differently defined inputs to estimate the efficiency of individual farms. Specifically, we use linear approximation of Hladik's DEA model [5] which was already used in efficiency analysis in agriculture [4].

Let n_1 is the number of inputs, n_2 the number of outputs and $m + 1$ the number of decision making units (DMU). Consider

- $\boldsymbol{I}_0 \in \mathbb{R}^{n_1}$ is the input nonnegative vector for DMU_0,
- $\boldsymbol{O}_0 \in \mathbb{R}^{n_2}$ is the output nonnegative vector for DMU_0,
- $\boldsymbol{I} \in \mathbb{R}^{m \times n_1}$ is the input nonnegative matrix for the other DMUs,
- $\boldsymbol{O} \in \mathbb{R}^{m \times n_2}$ is the output nonnegative matrix for the other DMUs.

Hladík's model uses the efficiency scale from 0 to 2, in which the higher scores (> 1) of production unit means that the unit remains efficient for larger variations of all data. The borderline for efficient unit is equal to 1. Lower scores (< 1) can be interpreted as the unit would be inefficient for larger variations of all data. Hence, the score is based on the largest allowable variation of all input and output data such that unit remains efficient or the smallest variation of data to become efficient in case of inefficient unit [5].

The linear model is as follows

$$\begin{aligned} \theta^* &= \max_{\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}} \theta \\ \text{s.t.} \quad \boldsymbol{O}_0^T \tilde{\boldsymbol{u}} &\geq 1 + \theta, \\ \boldsymbol{I}_0^T \tilde{\boldsymbol{v}} &\leq 1 - \theta, \\ \boldsymbol{O}\tilde{\boldsymbol{u}} - \boldsymbol{I}\tilde{\boldsymbol{v}} &\leq 0, \\ \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} &\geq 0, \end{aligned} \tag{1}$$

where $r = 1 + \theta^*$ is the resulting efficiency score of chosen DMU_0. As stated above, if $r \in (0, 1)$ then the DMU_0 is inefficient and if $r \in [1, 2)$ indicates that DMU_0 is efficient. In order to extract vectors of input and output weight, we can compute $\tilde{\boldsymbol{u}} := \boldsymbol{u}/(1 - \theta)$ and $\tilde{\boldsymbol{v}} := \boldsymbol{v}/(1 - \theta)$ with $\boldsymbol{u}$ and $\boldsymbol{v}$ represent the vectors of input and output weights, respectively.

The choice of inputs (and outputs) and their units of measurement is taken from [11]. The inputs are shown in Table 1. Eighteen different models are successively estimated, differing in the units of measurement of each input. Note that within groups the correlation between differently measured input is very high, above 0.9.

DEA Input	Description
Labour	L_1 – Total labour in annual working unit (AWU)
	L_2 – Total labour in wages in Euro
Capital	C_1 – Interest and depreciation in Euro
	C_2 – Total assets in Euro
	C_3 – Total fixed assets in Euro
Other input	O_1 – Total farming overheads and specific costs in Euro
	O_2 – Total specific costs in Euro
	O_3 – *no other input*
Utilised Land	A_1 – Utilised land in ha

Table 1 Variants of inputs for the DEA model.

For all combinations of inputs we use the same set of outputs. We consider the following variables as outputs:

1. Total output crops and crop production in Euro,
2. Total output livestock and livestock products in Euro,
3. Other output in Euro.

We estimate farm efficiency across all four states, separately for each year.

2.3 Regression Analysis

In the second stage, we estimate the effect of the selected variables on the efficiency estimated in the first stage. We assume following process for the second stage [9]

$$\theta_i = \psi(z_i^T \boldsymbol{\beta}) + \epsilon_i \geq 1, \tag{2}$$

where θ is output efficiency measure, ψ is smooth, continuous, function of vector regressors z_i and parameters $\boldsymbol{\beta}$ and $\epsilon \sim N(0, \sigma_\epsilon^2)$ is random variable with left-truncation at $1 - \psi(z_i\boldsymbol{\beta})$ representing the part of inefficiency, not explained by z_i. Because the efficiency of θ is unobservable, the form of the regression equation in the second stage is replaced by the first stage estimate of $\hat{\theta}$:

$$\hat{\theta}_i = \psi(z_i^T \boldsymbol{\beta}) + v_i \geq 1, \tag{3}$$

where $\hat{\theta}$ represents a first-stage efficiency estimate, v_i is serially correlated random error. Moreover v_i is correlated with z_i – more details about the asymptotic behavior of the random component v_i can be found in [8].

Simar and Wilson [8] derive assumptions when the second stage of the analysis leads to a truncated regression and propose Logit (or Tobit) model to estimate the truncated regression. Currently, Logit and Tobit regressions are the most commonly used methods for estimating the effect of variables on DEA efficiency [7].

The eighteen estimated DEA efficiencies ($\hat{\theta}$) in the first stage based on the combinations of inputs measurements from the Table 1 are used as the dependent variable in the second stage of the regression equation:

$$\hat{\theta}_i = \beta_0 + \beta_1 Sub_i + \beta_2 Sup_i + \boldsymbol{Year}_i^T \boldsymbol{\gamma} + \boldsymbol{Region}_i^T \boldsymbol{\omega} + \boldsymbol{Size}_i^T \boldsymbol{\alpha} + u_i, \tag{4}$$

where $i = 1, \ldots, 995$, the variable Sub_i represents total subsidies, Sup_i represents Total support, $\boldsymbol{Year_i}$ is a vector of dummy variables controlling for years 2014–2019 ($\boldsymbol{\gamma}$ is also a vector of estimated coefficients), u_i is a random variable. To reduce the likelihood of omitted variable bias, the model is augmented with additional control variables $\boldsymbol{Region_i}$ which is a vector of dummy variables to control for region fixed effect, $\boldsymbol{Size_i}$ is a vector of dummy variables to control for farm size.

The estimation of the equation 4 is done using logistic regression

$$P(\hat{\theta}_i = 1|z_i) = F_L(z_i'\boldsymbol{\phi}), \tag{5}$$

where F_L is distribution function for logistic distribution, $z_i = (Sub_i, Sup_i, \boldsymbol{Year}_i, \boldsymbol{Region}_i, \boldsymbol{Size}_i)$. The estimation of Eq. (5) is performed using the maximum likelihood method and the estimation of the variance-covariance matrix is performed using Bootstrap according to [8], specifically Algorithm 2.

However, the literature still does not provide an answer to the question of how to capture the dynamic behaviour of the efficiency panel. There are some efforts to dynamise the DEA method, but the statistical properties of this estimator are not yet known, nor has a consistent estimator for second-stage regression analyses been derived yet. For this reason, both first-stage and second-stage data are treated as cross-sectional data. Efficiency is estimated for each year separately.

3 Results and Conclusion

Figure 1 shows the estimates of the parameter β_1 from the equation 4 and Figure 2 shows the estimates of the parameter β_2 from the equation 4. The description of used inputs in each model is in Table 1. For both variables, the statistically significant parameter β_1 is evident for all eighteen DEA models. At the same time, the sign is positive for all eighteen DEA efficiencies. However, for example, DEA efficiency model $L_2C_1O_3$ is associated with the estimate $\hat{\beta}_1 = 0.004$ and DEA efficiency model $L_1C_3O_1$ with the estimate $\hat{\beta}_1 = 0.0021$. In the case of

FADN defines Total subsidies by code SE605
FADN defines Total support for rural development by code SE624
FADN defines denotes size by code SIZ6 and Economic size class (6 classes)

the parameter estimate β_2, we cannot reject H_0: $\beta_2 = 0$ for efficiency models $L_2C_1O_2$, $L_2C_2O_2$, and $L_2C_3O_2$. In other cases, the data support a negative and statistically significant effect of Total Support on efficiency. As with Total Subsidy, we see considerable variability in the estimates of β_2. For example, DEA efficiency model $L_1C_1O_2$ is associated with the estimate $\hat{\beta}_2 = -0.076$ and further DEA efficiency model $L_2C_2O_1$ is associated with the estimate $\hat{\beta}_2 = -0.0121$.

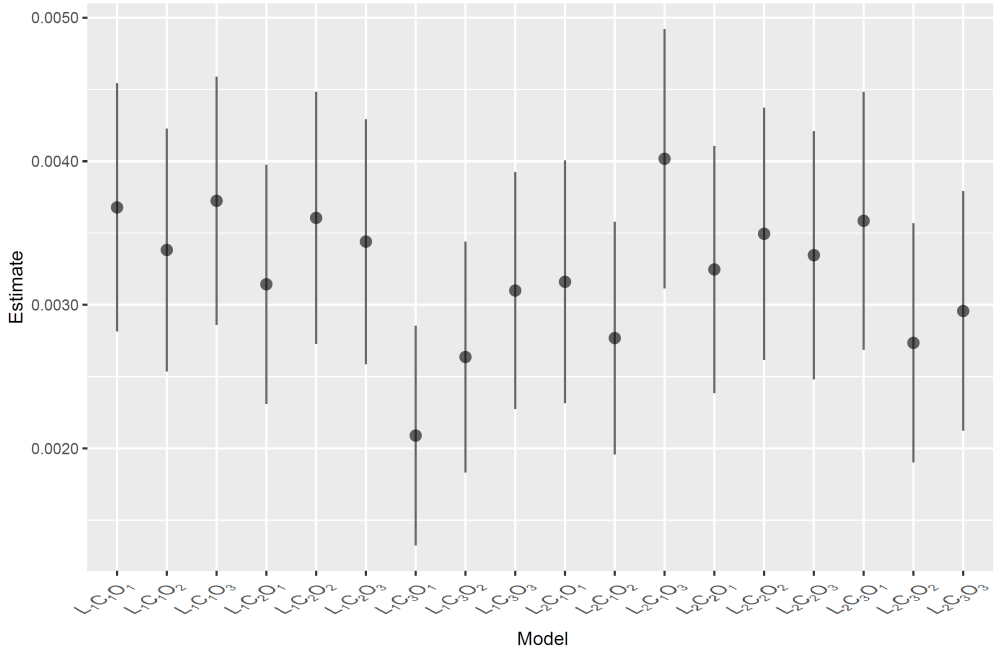


Figure 1 Total subsidies – the description of the efficiency is in Table 1.

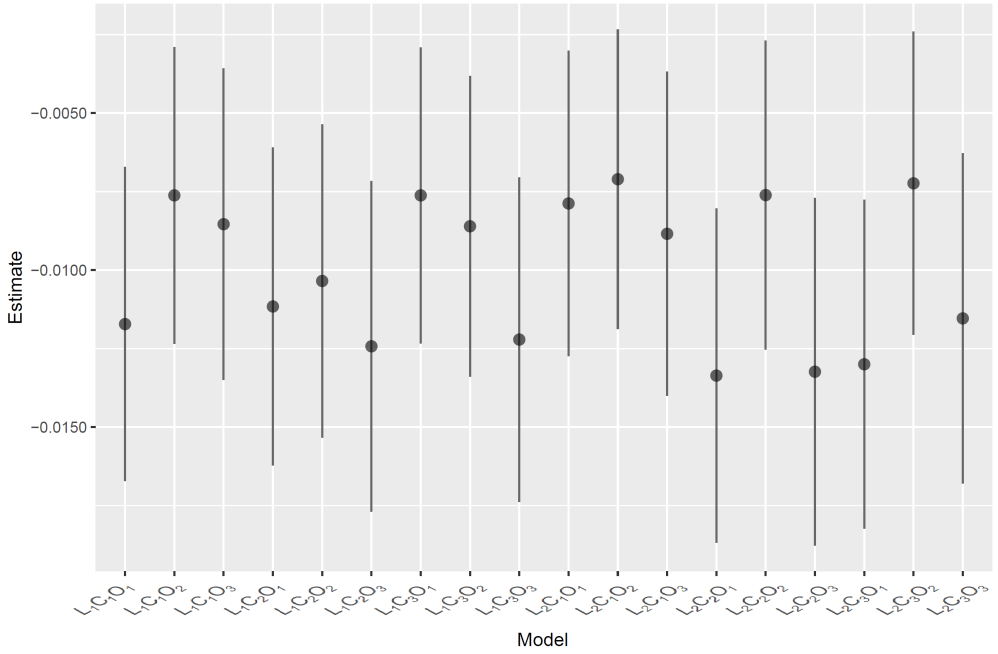


Figure 2 Total support – the description of the efficiency is in Table 1.

It is clear from the regression analysis that different approximations of labour, land, capital and additional inputs lead to different conclusions in the second stage of the regression analysis. This problem is particularly acute for the estimation of the parameter β_2. In this case, there is considerable heterogeneity in the estimates and in three cases the parameter β_2 appears statistically insignificant. It is clear from the regression analysis that different approximations of labour, land, capital and additional inputs lead to different conclusions in the second stage of the

regression analysis. This problem is particularly acute for the estimation of the parameter β_2. In this case, there is considerable heterogeneity in the estimates and in three cases the parameter β_2 appears statistically insignificant. On the one hand, this phenomenon may explain, for example, the considerable variability in the results of empirical analyses focusing on the effect of subsidies on efficiency, see, for example, [7]. However, it is misleading to compare the results of different papers in the case of differently approximated inputs. At the same time, emphasis should be placed on the interpretation of the regression parameters. These parameters need to be viewed in the light of the effect on the specific technical efficiency that is given by the inputs used, and not to generalize conclusions without considering that different approximations of the inputs lead to different efficiency sets and hence different technical efficiencies.

Acknowledgements

The work was supported by the Internal Grant Agency of Prague University of Economics and Business under Grant F4/24/2023.

References

[1] Banker, R. D., Charnes, A. & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science* 30, 1078–1092.

[2] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research* 2, 429–444.

[3] Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S. & Shale, E. A. (2001). Pitfalls and protocols in dea. *European Journal of operational research* 132, 245–259.

[4] Frýd, L. & Sokol, O. (2021). Relationships between technical efficiency and subsidies for czech farms: A two-stage robust approach. *Socio-Economic Planning Sciences* 78, 101059.

[5] Hladík, M. (2019). Universal efficiency scores in data envelopment analysis based on a robust approach. *Expert Systems with Applications* 122, 242–252.

[6] Kneip, A., Simar, L. & Wilson, P. W. (2015). When bias kills the variance: Central limit theorems for dea and fdh efficiency scores. *Econometric Theory* 31, 394–422.

[7] Minviel, J. J. & Latruffe, L. (2017). Effect of public subsidies on farm technical efficiency: a meta-analysis of empirical results. *Applied Economics* 49, 213–226.

[8] Simar, L. & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics* 136, 31–64.

[9] Simar, L. & Wilson, P. W. (2011). Two-stage dea: caveat emptor. *Journal of Productivity Analysis* 36 , 205–218.

[10] Simar, L. & Zelenyuk, V. (2020). Improving finite sample approximation by central limit theorems for estimates from data envelopment analysis. *European Journal of Operational Research* 284, 1002–1015.

[11] Sokol, O. & Frýd, L. (2023). Dea efficiency in agriculture: Measurement unit issues. *Socio-Economic Planning Sciences* 86, 101497.

Economic Perspective of Smart System for Waste Collection

Taťána Funioková[1], Petr Kozel[2], František Zapletal[3]

Abstract. Smart cities use modern information technologies to improve the quality of life of their residents and to become attractive to other entities as well. One of the areas that is essential for every municipality and where there is room for the application of modern approaches is waste collection. This article aims to extend the concept presented by [11]. Using mathematical programming methods in combination with financial analysis, a thorough analysis of the viability of the smart system is carried out with regard to fluctuations in requirements during the year. Next, an economic analysis is performed that compares the smart system with a traditional system with fixed collection. The results showed that the smart system would impact each resident groups differently: the costs for companies and couples will increase, on the other hand, the companies would reach substantially greater level of comfort. For most families, the costs would be reduced and the comfort increased.

Keywords: Smart city, waste collection, optimisation, linear programming.

JEL classification: C44
AMS classification: 90C05

1 Introduction

The smart city concept is the optimisation of related processes with respect to the costs, ecology and quality of life of citizens through smart technologies and data analysis [2]. However, the implementation of new solutions is often associated with high initial costs, which may be unaffordable for small towns. On the other hand, it is a challenge to find new smart solutions that are cheap and easy to use.

Usually, a fixed schedule for waste collection is applied in cities. It means that their locations have been served repeatedly on predetermined days. Such an approach is easy to manage for a waste company, but it is not necessarily optimal for residents. The amount of waste produced each day is a random variable and depends on many factors. It can easily happen that a place is serviced when the bin is almost empty, on the other hand, one can waste too long for serving.

The smart systems of waste collection, which can be found in the literature, require new hardware devices. Namely, many smart models can be found in the literature using smart devices and sensors, see [6], [9], [7], [10], or [1]. These sensors can measure the volume or weight of a bin. The model introduced by [3] is an exception – people report the need to serve bins in public space. This article deals with smart waste collection for residents that does not require any new hardware equipment. Namely, we investigate the system proposed by [11]. In this case, we assume the cooperation of residents who report the filling status of their bins themselves through a mobile or web application. The authors of [11] concluded that the smart system is potentially more comfortable for residents, despite the fact that it brings a slightly greater distance travelled. However, the initial study does not include a detailed sensitivity analysis and assessment of the economic impact on affected residents. This study follows, in particular, the following goals:

- check the viability of the smart system proposed by the authors recently [11];
- investigate the impact of the smart system on the various resident groups;
- investigate the impact of the shocks in the waste production during year;
- propose the economic settings of the smart system;

All analyses are performed using the same network with 50 residential nodes (and one depot) as it was used in [11] to guarantee better comparability of the results. To be able to assess the real impact of the new system on various types of residents, who differ in the number of people living in one place, the waste production rate, and the yearly fee paid for the waste collection within the current system too. A thorough sensitivity analysis is done to check the robustness of the results.

The remainder of this paper is organised as follows. Sec. 2 provides a brief reminder of the model proposed by [11] and its assumptions. Sec. 3 contains the input data. The most important section is Sec. 4 where the results of

[1]VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, tatana.funiokova@vsb.cz
[2]VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, petr.kozel@vsb.cz
[3]VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, frantisek.zapletal@vsb.cz

the analysis and their robustness are checked. The last section (Sec. 5) summarised the obtained results.

2 Smart Model and its Assumptions

Due to the fact that the model has been carefully mathematically described in [11], only the main logic of this model is recalled here. An interested reader can look into the initial study.

The basic idea of the smart model is to replace the fixed serving schedule with a flexible one based on serving on the request of residents. Namely, each resident is expected to provide the report (e.g., using some smart application, SMS or websites) that his or her bin is full and needs serving. Besides that, to smooth out fluctuations, a similar report is done when the bin is half-full (i.e., the state when the resident is convinced that about half of space in the bin is used). Thus, the model is based on two-stage reporting.

Each bin can be served anytime after the first report (about the half-full state) is done, but not later than on the d-th day after the second report. This setting should bring sufficient comfort to residents, but also the flexibility of the system.

All parameters of the model are considered fixed and non-random, except for the waste production rates. These rates are a crucial factor based on which the quality of the model is assessed. Namely, according to [11], several groups of residents are distinguished. Each group is defined by its random waste production rate (and for further economic analysis, by the number of people living in the given node too).

The whole algorithm of the model can be summarized in the following steps:
1. find the status of each node (half-full, full since that day, full and waiting for the first day, full and waiting the second day);
2. determine which nodes have to be served that day;
3. find an optimal route for serving the nodes identified in the previous step.

Within the optimisation process, more mathematical methods are used, depending on the usage of the truck capacity c. Particular cases will be briefly described. The aim of optimisation is to minimise the total distance travelled by serving.

The simplest case occurs when the following equation holds:

$$\sum_{j \in I} b_j = c,$$

where I denotes the set of nodes (residents) that must be served in the given step (i.e., the nodes waiting the second day after reporting the full bin), c stands for the truck capacity, and b_j is the request of the j -th node. In this case, the shortest path, starting and ending at the depot and going through all nodes in I, must be found. Depending on the size of the problem, one of the known exact or heuristic algorithms for finding the minimum Hamiltonian cycle can be used, see [8].

If there is still some unused capacity of the truck after serving all nodes in I (i.e., if $c - \sum_{j \in I} b_j > 0$), then this capacity will be used to serve the nodes from the set of nodes J, which can possibly be served that day (i.e., the nodes which already reported either a full bin, or at least a half-full bin). Primarily, full bins are served. If not all nodes with full bins can be served because of the capacity, the nodes with the optimal location with respect to the nodes in I will be chosen. These nodes can be identified by solving the mathematical model derived from the assignment optimisation problem. The goal is to assign the given number of nodes in J to the nodes in I, respecting the objective of minimising the distance. Unlike a classic assignment problem, the assignment constraints are modified as follows. Each node in J can be assigned to a single node in I only. However, any number of nodes in J can be assigned to each node in I. A more detailed description of possible interactions between I and J can be found in [11].

For the sake of completeness, it can happen that the number of nodes in I exceeds the capacity of the truck c (or its p multiple), i.e., $\sum_{j \in I} b_j \geq p \cdot c$. In this case, it is necessary to include also some elements of decomposition in the optimisation process, e.g., using the Cluster-First Route-Second method. Such methods are described in detail in [5] and [4].

Since each model is just a simplification of reality, one must be very careful about what assumptions must be met to make the results reliable. Most of the assumptions are identical to those declared in [11]. However, some additional ones had to be adopted to also perform the economic and environmental analysis.

The performance of the smart system will be compared with the current state. This current state corresponds to

the model which is applied very often in villages nowadays. Namely, the fixed route is realised every two weeks, i.e., all nodes are served every 14 days regardless of whether they do need serving or not, and whether some shock (extraordinary days in terms of waste production rate) occurred or not. The current system is funded according to the number of people living at each node (and companies pay an amount according to the local directives).

From the perspective of residents, the following simplifying assumptions are set:
- each node is equal to one bin to be served (no node has more than 1 bin);
- residents responsibly report the usage status of their bins. In particular, two states are distinguished: "a bin is approximately half-full" and "a bin is full and needs serving";
- each resident cannot wait more than 2 days for serving after the report of full usage was done;
- each node can be served anytime after reporting the half-full status.

From the perspective of policy-makers, other assumptions must be added:
- the truck capacity is equal to c (i.e., c full bins, or $2c$ half-full bins, or their combinations);
- the capacity of a bin reported as half-full is considered half-full until it is reported as full;
- if it is not possible to serve all the nodes which have to be served that day, the truck must go (at least) 2 routes that day;
- the waste production follows the uniform probability distributions with different parameters for each resident group;
- only the volume of waste does matter (regardless its weight or type).

3 Input Data

This section contains a description of the input data used for verification, as well as all considered settings.

The model will be verified using a randomly generated network with 50 residents (serving nodes) and a single depot that is both the starting and the terminating point of the truck. The same network was used as in the case of [11]. The 50 resident nodes were randomly split into 5 resident groups (the proportion of residents in each group reflects an average expected structure). Namely, pensioners (living as singles or in pairs, with a very low waste production rate), couples (productive adults without children), small families (with a single child), big families (with 2 or 3 children) and companies (producing the greatest volume of waste) are distinguished. The numbers of residents in each group and their properties (probability distribution of the waste production, and number of people living together in the node) can be found in Tab. 1. For the sake of simplicity, the waste production considers a uniform probability distribution given by minimum and maximum values for all groups.

Group name	Code	Number of nodes	Minimum [days]	Maximum [days]	Number of people
Pensioners	1	13	10	17	1 or 2
Couples	2	8	8	14	2
Small families	3	17	6	11	3
Big families	4	6	4	8	4 or 5
Companies	5	6	2	5	not applicable

Table 1 Parameters of uniform probability distributions for each group of residents [11]

The model will be run for the bin capacity c equal to 6 to be able to compare the results with those presented by [11]. The period of 92 days (1/4 of a year) has been considered. The distances between the nodes are known. In the initial state (at $t = 0$) all bins are completely empty. The results of all analyses for the smart system and the fixed system with service every 14 days (denoted as F14) will be compared. The minimum distance travelled under F14 is equal to 880 km, see the results in [11].

To perform the economic analysis, the quarterly fee under the current fixed system with a period of 14 days has to be known. This value is set to 200 CZK per single person in the node (the company has to pay 1000 CZK).

The detailed results (by nodes) will be provided only for a single run of the model (for the whole time period of 92 days). Because the results depend also on the outcomes of the random variables (waste production), the robustness check will be done using 500 runs, for which the mean value and standard deviations will be calculated.

It is reasonable to expect that the intelligent system will be more resistant to external shocks to waste production rates caused by e.g. special times of the year such as Christmas, school holidays, etc. Therefore, the results of the model without shocks (NS) and with shocks (WS) will be also compared. For this purpose, we model the non-standard waste production WS using a scenario with two 8-day shocks. The first of these, representing increased waste production, was initiated on day 30 ($t = 30, \ldots, 37$). During this period, all units produced twice

the normal amount of waste. The second shock, representing reduced waste production, was deployed on day 62 ($t = 62, \ldots, 69$). During this period, all units produced half the normal amount of waste.

In addition, companies can be considered a very special population group, as they have the highest waste production rates. This group also has a special fee regime (the fee is not set based on the number of people in the node), and the effect of the shocks on waste generation can be expected to differ from other groups; e.g. during the summer months most residents produce less amount of waste due to vacations, travelling, but the companies may behave without any change. For this reason, the setting where shocks do not affect the companies will also be distinguished (that is, the shocks are ignored for the companies). This setting will be referred to as WSC.

4 Results of the Analyses

First of all, the impact of the smart system on the comfort for the residents is investigated. Comfort is given by the time to wait for serving when the bin is full. Let us start with the issue of whether the smart system provides on average a better or worse level of comfort. For a single run, the results showed that 29 of 50 nodes improved their comfort when shocks were included (WS). For the setting without shocks (NS), the comfort was increased for 27 nodes. To quantify these changes, the results obtained from all 500 runs were used. Three scenarios are considered: the scenario (S_M) with mean values for all 500 runs, the pessimistic scenario (S_P) where the differences between the best possible value of the fixed system and the worst possible value of the smart system were adopted, and the optimistic scenario (S_O) where we considered the difference between the worst possible value in the fixed system and the best possible value in the smart system. The resulting values, revealing by how many days of waiting the fixed system (F14) is better than the smart model (WS), can be found in Tab. 2. The values correspond to the average values per each node in a resident group for whole 92 days. Negative values indicate better performance of the smart system and vice versa. It can be seen that improvement in comfort is expected for families and companies. On the other hand, a small decrease in comfort can be expected for pensioners and couples. However, a decrease of between 1 and 2 days over the whole period should be acceptable for any group, given the smart system setup. Shocks enforce the benefit of the smart system. The reason is that the smart system can react more flexibly to shocks. It can be concluded that, in terms of comfort, the smart system is the most beneficial for large families and companies, and that shocks in the waste production rate favour the smart system even more.

Group	S_O (WS)	S_M (WS)	S_P (WS)	S_O (NS)	S_M (NS)	S_P (NS)
1	-0.31	1.77	3.15	0.77	1.86	3.38
2	-3.88	1.65	4.38	1.38	2.94	4.38
3	-15.53	-1.80	5.00	-9.41	3.25	5.18
4	-33.67	-14.23	2.83	-32.17	-7.01	8.33
5	-49.83	-39.93	-1.17	-53.00	-32.10	-12.00

Table 2 Comparison of average waiting time of a node between the fixed and the smart system

Now, let us focus on the economic point of view. Namely, it will be investigated how the fee per single serving must be set to collect the same amount of money as in the current fixed (F14) system. It must be taken into account that the total number of kilometres under the smart system is slightly greater than under the current fixed system. This result was also confirmed by [11]. The optimal distance travelled in the fixed system is 880 km (see [11] for further details. The smart system without the shocks (NS) resulted in 918.9 km, 979.4 with the shocks (WS) and 971.3 under the WSC setting (the mean values of 500 repeated runs equal to 925 km (NS) and 933 km (WS). It means that the smart system brings approximately 5% greater distance travelled, which can be regarded as cost of the flexibility of the system and improved comfort. Therefore, it is assumed that the total amount of money to be collected is 5% greater than in case of the fixed system, which, according to the settings explained in Sec. 3, operates with 27,900 CZK. To obtain the fee per node and a single serving of a full bin, the total number of serving across all 50 nodes was used. The fee to serve a half-full container is half of the full fee. The full fee is equal to 148 CZK/serving (i.e., 74 CZK for serving the half-full bin). The results of a single run showed that 18 nodes out of 50 would pay less in the smart system than in the fixed system (F14) when no shocks occurred (NS), exactly 50% nodes paid less if the shocks were considered (WS) and 27 out of 50 nodes paid less when the shocks were not applied to the companies (WSC). Again, when looking at the results of 500 runs, 26 nodes out of 50 paid less in the smart system.

When exploring the results of the economic analysis in more detail, the average change of the fee for a node in a group can be calculated, see Tab. 3. Negative values indicate lower costs of the smart model and vice versa. The results there were calculated based on a single run. It can be seen that families can expect lower costs and higher costs by other stakeholder groups. When checking the robustness using 500 runs, the results do not differ much

– the mean values for the setting with shocks (WS) are provided in Tab. 4. Negative values indicate lower costs of the smart model and vice versa. The greatest absolute increase in costs can be observed for companies. On the other hand, this increase is only around 20% and is balanced by increased comfort. The most troubles are caused to the couples and the greatest benefit is brought to large families. An interested reader can look at the Appendix where the precise values of costs, calculated within a single run, are provided for each single node.

Group	Mean value of change (WS) [CZK]	Mean value of change (NS) [CZK]
1	62.7	36.1
2	153.7	154.5
3	-7.4	-29.6
4	-192.6	-125.3
5	172.8	224.7

Table 3 AVG change of the costs of a node between the fixed (F14) and smart system [CZK]

Group	Mean value of the change (WS) [CZK]
1	48.0
2	141.5
3	-24.9
4	-138.8
5	216.7

Table 4 AVG change of the costs of a node between the fixed (F14) and smart system including shocks (WS)

The results for both criteria (comfort and costs) are concluded in Tab. 5. Namely, the number of nodes (distinguished according to stakeholder groups) which would improve/decrease their comfort/costs for a single run can be found there. It can be seen that 16 out of 50 nodes improved their performance in terms of both criteria (1 pensioner, 10 small families and 5 large families). In contrast, 12 nodes decreased their performance in both criteria (5 pensioners and 7 couples).

Group	Count	1	2	3	4	5
Better comfort and lower costs	16	1	0	10	5	0
Worse comfort and lower costs	9	5	1	3	0	0
Better comfort and higher costs	13	2	0	4	1	6
Worse comfort and higher costs	12	5	7	0	0	0

Table 5 Overall results of the analyses in terms of comfort and costs for a single run

It can be seen that not all residents would welcome the smart system. On the other hand, without any doubt, it would be more fair because it is independent of the number of people living in place and depends just on the quantity of waste produced. This should ideally cause pressure on the waste reduction and greater proportion of sorted waste. Moreover, if all residents reduce their waste, the number of kilometres travelled would also decrease, and the fee per a single serving would be lower.

5 Conclusions

This paper presents a deep critical analysis of the smart system of waste collection proposed by [11]. The original study missed the robustness analysis and absolutely ignored that the volume of waste produced during the year is not constant. Moreover, any analysis of the economic viability of the proposed system was missing too. All these drawbacks were addressed and investigated in this paper.

The results showed that the new system, under the given assumptions, leads to a greater travelled distance (approximately by 5%), which requires more money to cover the increased costs. On the other hand, for selected resident groups (families and companies), it brought better comfort (lower waiting time for serving). As for the economic perspective, when considering the fixed fee for each serving, the smart system brings lower costs, mainly for families. On the contrary, companies and couples without children would pay more. Although some residents (about one quarter) would be harmed by the new system (they would have lower comfort and higher fee), it should be noted that the proposed system is 1) more fair since the costs depend only on the quantity of the waste produced,

and 2) more resistant against shocks in the waste production. It is also worth noting that the new system should motivate people to reduce the waste volume more than the current system (the fee should be recalculated based on the number of travelled kilometres, which is dependent on the number of requests).

Further research should focus on relaxing selected assumptions and further analysis of shocks.

Acknowledgements

This work was supported by the SGS project No. SP2023/078. This support is gratefully acknowledged.

References

[1] Gutierrez, J. M., Jensen, M., Henius, M. & Riaz, T. (2015). Smart waste collection system based on location intelligence. *Procedia Computer Science*, 61, 120–127.

[2] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J.,Paraszczak, J. & Williams, P. (2010). Foundations for smarter cities. *IBM Journal of Research and Development*, 54(4), 1–16.

[3] Kalpana, M. & Jayachitra, J. (2017). Intelligent bin management system for smart city using mobile application. *Asian Journal of Applied Science and Technology (AJAST)*, 1(5), 172-175.

[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*, 252–257.

[5] 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*, 584–589.

[6] Lu, J. W., Chang, N. B., Liao, L. & Liao, M. Y. (2015). Smart and green urban solid waste collection systems: advances, challenges, and perspectives. *IEEE Systems Journal*, 11(4), 2804–2817.

[7] Popa, C. L., Carutasu, G., Cotet, C. E., Carutasu, N. L. & Dobrescu, T. (2017). Smart city platform development for an automated waste collection system. *Sustainability*, 9(11), 2064.

[8] Rahman, M. S. & Kaykobad, M. (2005). On Hamiltonian cycles and Hamiltonian paths. *Information Processing Letters*, 94(1), 37-41.

[9] Ramos, T. R. P., de Morais, C. S. & Barbosa-Póvoa, A. P. (2018). The smart waste collection routing problem: Alternative operational management approaches. *Expert Systems with Applications*, 103, 146–158.

[10] Soh, Z. H. C., Husa, M. A. A. H., Abdullah, S. A. C. & Shafie, M. A. (2019). Smart waste collection monitoring and alert system via IoT. *In 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 50–54.

[11] Zapletal, F., Kozel, P. & Chytilová, L. (2022). Using genetic algorithm for advanced municipal waste collection in smart city. *In Proceedings of Mathematical methods in Economics conference 2022, Jihlava, September 7th – 9th, 2022*, 411–417.

Appendix

ID	No. of full	No. of half-full	Fee (WS) [CZK]	Fee (F14) [CZK]	Difference [CZK]	Group
1	4	0	593	600	-7	3
2	2	1	370	200	170	1
3	4	2	741	800	-59	4
4	1	3	370	200	170	1
5	3	2	593	600	-7	3
6	3	0	444	400	44	2
7	3	1	519	600	-81	3
8	3	0	444	200	244	2
9	5	0	741	1000	-259	4
10	7	0	1037	1000	37	5
11	7	1	1111	1000	111	5
12	3	1	519	600	-81	3
13	3	0	444	400	44	2
14	3	0	444	200	244	2
15	2	1	370	400	-30	1
16	4	1	667	600	67	3
17	3	3	667	600	67	3
18	7	3	1259	1000	259	5
19	2	2	444	200	244	1
20	6	4	1185	1000	185	5
21	1	6	593	600	-7	3
22	7	3	1259	1000	259	5
23	1	2	296	200	96	1
24	3	1	519	600	-81	3
25	4	0	593	600	-7	3
26	4	2	741	600	141	3
27	3	2	593	600	-7	3
28	2	0	296	400	-104	1
29	3	2	593	1000	-407	4
30	1	3	370	400	-30	2
31	3	1	519	400	119	2
32	8	0	1185	1000	185	5
33	3	4	741	600	141	3
34	1	3	370	200	170	1
35	1	5	519	600	-81	3
36	1	3	370	400	-30	1
37	2	1	370	400	-30	1
38	3	5	815	800	15	4
39	1	3	370	200	170	1
40	2	3	519	600	-81	3
41	3	1	519	600	-81	3
42	3	1	519	200	319	2
43	2	1	370	400	-30	1
44	2	2	444	400	44	1
45	2	2	444	200	244	2
46	3	2	593	600	-7	3
47	3	2	593	600	-7	3
48	0	5	370	400	-30	1
49	4	3	815	1000	-185	4
50	4	2	741	1000	-259	4

Table 6 Detailed comparison of the fees under the fixed system (F14) and the smart system (WS)

DEA Methodology as a Tool for Determining the Efficiency of Public Transport in South Moravian Region

Nino Gochitidze [1]

Abstract. This paper discusses the possibility to apply DEA methodology for determining the efficiency of public transport. There are 673 municipalities in the region connected by municipal and private transport companies in a unified mode. According to the latest data, transported passengers among region have increased from 118 to 828 million of passengers-kilometers during the period 2004–2019, while in the city of Brno from 344 to 403 million passengers. Along with the increase in the flow of passengers, the criteria for efficiency evaluation also changes.

The aim of this contribution is to (i) identify the main evaluation criteria of transport efficiency for passengers based on a pilot survey in the region, (ii) identify a set of feasible transportation alternatives, and (iii) by using Data envelopment analysis (DEA), find an efficient transport design and discuss the suitability of the method for the transportation problem.
Keywords: Public Transport, South Moravian Region, DEA, Efficiency.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Public transport as a part of worldwide logistics nowadays combines several types of transports, provided by one or several private transport companies, as well as by transit authorities. Readily apparent that increase of population itself dictate establish extra criteria's concerning urbanization and transport intensity increase. By assumption of European Commission's White Paper [5] intensity of passenger's transport will increase approximately 34% by 2030, compare to 2005. Hence, it is necessary to take innovative measures to halt the progression of the mentioned events by increasing the efficiency of the existing transport modes and their combinations. Efficiency can be understood as the highest level of performance that uses the minimum amount of inputs to obtain the maximum amount of output. In other words, the more efficient the process or unit is, the fewer inputs are needed to produce the outputs. Changes in city transport management are a complex process in order to implement the correct and most suitable transport strategy. Once developed new areas must meet the demands and needs of all possible stakeholders (population, government, transport providers, etc.). It's even harder to evaluate transport process efficiency due to the fact that every decision-maker has different evaluation criteria. In the South Moravian Region (SMR), the public transport system nowadays is combined and named the Integrated Transport System (IDS JMK). The system unifies all modes of transport, and they are tightly linked to each other. This is evidenced by annual passenger growth, which indirectly can be indicated as environmental benefit as well as an indication of passengers' satisfaction with public transport. More specifically, saving resources and reducing individual car traffic. In preliminary research to employ models with both DEA basic output-oriented methods BCC and CCR, compare their results ; once in studies [6], [7] models showed quite representative result. The aim of this contribution is to (i) identify the main evaluation criteria of transport efficiency for passengers based on a pilot survey in the region, (ii) identify the set of feasible transportation alternatives, and (iii) by using Data envelopment analysis (DEA), find an efficient transport design.

2 Materials and Methods

The final aim of the research is to compare different transport combinations and find efficient transport designs. The total of eight possible combinations of transportation were estimated by Data Envelopment Analysis (DEA). The evaluation of transport combinations was built on the following indicators: vehicle carbon footprint, travel time, travel cost, and frequency of used transport combinations. Each piece of data was calculated for a specific

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1665/1 Brno, 613 00, Czech Republic, xgochiti@mendelu.cz; nino.gochitidze@gmail.com;

vehicle, and average amount was chosen. The DEA method is based on linear programming and determines dependencies between inputs and outputs, which was introduced by Charnes, Cooper, and Rhodes in late 20 [th] century [3]. The DEA approach has a non-parametric character, which allows of its use without knowledge of functional dependencies between outputs and inputs (possibility of using different units of measure). The method comes up along with several additional opportunities for use, practically collaborating with decision-makers and analysts, which evolves as a result of cooperation in input and output choices to be used and consists of selecting the hypothetical situation categories. Final efficiency scores generally lie within the interval. The value assigned to the decision unit from the interval depends not only on the other units in the model but also on the inputs and outputs of the unit. However, this method allows to change the number of inputs to measure performance efficiency as accurately as possible. Two basic models of the DEA method are identified: (1) CCR (Charnes, Cooper, and Rhodes) [3] – model hypothesises constant returns to scale (CRS). (2) BCC (Banker, Charnes, and Cooper) – model hypothesises variable returns to scales (VRS) [2].

Each of the DEA's basic models CCR and BCC are classified as either input-oriented (I-O) or output-oriented (O-O). Both models are different in their specifications; for example, I-O models aim to produce at least the perceived output while minimizing the inputs, whereas O-O model maximizes its outputs using the most measured levels of input. DEA methodology is applied in different areas, such as health care, benchmarking, banking, supply chain, transportation, business economics and management [11], [17]. Methodology has been selected multiple times for evaluating the performance of the transportation system as in terms of technological as well as financial considerations. In most of the research in transportation fields, the output variables are chosen number of passengers, vehicle kilometers, and length of urban roads. In [13] , DEA methodology (input/output-oriented BCC models) is used for assessment of the safety and security of intermodal transportation facilities, where the number of passengers and cargo (tonnage) are chosen as output variables. Another example of DEA usage in the transportation field is [18], where the BCC-O model is applied for ridership maximization, using the following output variables: average passengers and vehicle kilometers per day. In [9], operational income and the number of daily passengers is chosen for calculation of operational and service efficiency based on the output-oriented BCC DEA model. In studies regarding public transport efficiency evaluation in the Czech Republic [6], [7], employees, energy and rolling stock are utilized as input variables and passengers' revenues output variable. For evaluation of efficiency, two-stage DEA CCR and BCC models are used. BCC model is applied in the study [8] for the efficiency evaluation of urban transport in different Polish cities. There are dozens of studies using different DEA models for transport-related operations, namely [4]: Stochastic Frontier Analysis; [1]: integration of traditional DEA by combination the Delphi technique with AHP method; Some authors have preferred combinations of the CCR and BCC models for comparisons and more transparency. Namely, in [15], authors used three-stage DEA model approach for evaluation of bus transport operations' efficiency. Based on above-mentioned researches and [10], choosing model orientation depends on which variable the decision maker has control over. In our scenario, the frequency of the transport combination is maintained by passengers. Accordingly, for transport combination efficiency evaluation in this study, basic output-oriented DEA models, particularly CCR-O and BCC-O, will be applied. If there is a set of DMU$_j$ (j = 1, ..., n); input vector $(x_{1j}, \ldots, x_{mj})$ and input weight vector $(v_1, \ldots, v_m)$ of DMU$_j$; output vector $(y_{1j}, \ldots, y_{qj})$ and output weight vector $(u_1, \ldots, u_q)$ of DMU$_j$. Assuming that DMU$_j$ consumes x_{ij} amount of input i to produce y_{rj} amount of output r, input and output of DMU$_k$ (k = 1, ..., n) being evaluated as $(x_{1k}, \ldots, x_{mk})$ and $(y_{1k}, \ldots, y_{qk})$, where $x_{ik} \geq 0$ and $y_{rk} \geq 0$. $\mu_r = tu_r$ and $v_i = tv_i$, where $t = (\sum_{i=1}^{m} v_i x_{ik})^{-1}$, μ_0 and v_0 are free variables. Accordingly output-oriented DEA CCR and BCC models can be expressed as follows:

Output-oriented CCR DEA model

$$min \frac{\sum_{i=1}^{m} v_i x_{ik}}{\sum_{r=1}^{q} u_r y_{rk}} \quad \text{subject to} \quad \frac{\sum_{i=1}^{m} v_i x_{ij}}{\sum_{r=1}^{q} u_r y_{rk}} \geq 1 \quad (j = 1, \ldots, n),$$

$$u_r \geq 0 \ (r = 1, \ldots, q), \quad v_i \geq 0 \ \ (i = 1, \ldots, m).$$

Output-oriented BCC DEA model

$$max \sum_{i=1}^{m} v_i x_{ik} + v_0 \ \text{ subject to } \sum_{r=1}^{q} \mu_r y_{rj-} \sum_{i=1}^{m} v_i x_{ij} + v_0 \leq 0 \ \ (j = 1, \ldots, n),$$

$$\sum_{r=1}^{q} \mu_r y_{rk} = 1, \quad \mu_r \geq 0 \ (r = 1, \ldots, q), \ v_i \geq 0 \ \ (i = 1, \ldots, m), \ v_0 \in \mathrm{R}.$$

Structure of the Questionnaire

For data gathering and to identify the main evaluation criteria of transport efficiency for passengers, the study utilized the survey method. In the pilot survey, target respondents were chosen according to their permanent residence and the frequency of daily movement among SMR. A total of 25 questions were included in the design of

the questionnaire. The structure of the questionnaire was inspired by [12], based on which demographic, multiple choice, single-select, Likert scale, matrix, and ranking types of questions were utilized. Despite the abundance of questions, the most attention will be paid to questions revealing the passengers' transportation habits, namely means of transport owned by households; regular travel destinations from permanent residence; reasons for not using public or regional transport; dependence between bicycle infrastructure improvement and the use of the bicycle as main transport; frequency of use individual or combination transport; importance level of transportation characteristics from passengers' personal perspective; As transportation is a massive source of carbon emissions [14], questions were used to reveal changes, that may be a reason for passengers (using own car as main transport) to switch to a relatively more environmental option; policies that may promote walking, bicycles, and other eco-friendly vehicles; Despite such a wide range of data, only the main research data will be presented in this article (Tables 1 and 2), in particular data useful for further efficiency evaluation of multiple transport combinations.

Interpretation of Main Collected Data

One of the paper aims is the identification of feasible alternatives for transportation. Accordingly, eight different combinations and regional/city travel frequencies were proposed for the surveyed passengers. The questionnaire has shown more clearly residents' movement directions and the most frequently used transport in the region (see Table 1). If we sum up results by the regular or daily frequency of the combinations, the most used arrangements are № 2, 3, 5, 6, 7, and 8; while on a weekly level, № 1, 2, 3, 4, 5, and 8. From here, we can assume that travelling by car is popular on a daily basis, while combinations of public transport are popular on a weekly or monthly basis. For numerical representation of the daily used combination frequency, the obtained data was assigned coefficients:

$$\text{never: } 0;\ \text{rarely: } 1/30;\ \text{often: } 1/4;\ \text{regularly: } 5/7;\ \text{always: } 1. \qquad (1)$$

The coefficients were adjusted to convert the given monthly and weekly frequencies to daily values (used combination by passenger per day).

	f_1	f_2	f_3	f_4	f_5
Transport combinations / Frequencies	**Never**	**Rarely** (1-3 times a month)	**Often** (1-3 times a week)	**Regularly** (4 - 5 times a week)	**Always** (almost every day)
	[%]	[%]	[%]	[%]	[%]
1. Regional Transport & Brno City Public Transport	25	40	35	0	0
2. Regional Transport & Walk	35	45	10	5	5
3. Regional Transport & Bicycle	70	25	5	5	0
4. Regional Transport & Shared Electric car/Car/Taxi	75	15	10	0	0
5. Own car & Brno City Public Transport	55	20	10	5	10
6. Own car & Shared Electric car/Car/Taxi	85	10	0	5	0
7. Own car & Bicycle	80	15	0	5	5
8. Own car & Walk/e-Bike/e-Scooter/ Hoover Boards etc.	60	20	10	5	5

Table 1 Frequency of transport combinations used by passengers

The result also has shown the detailed level of importance of each transport characteristic. According to the current answer matrix - Table 2, if the conditional efficiency margin is limited by first three levels of the significance, the most important transport characteristics for passengers are “on-time performance” /” adherence to the timetable”, “safety and security,” “speed of arrival,” “safety and security” and “price”. As relatively important characteristics: “punctuality”/ “a regular schedule”, “passengers waiting time” and “quality” / “reliability”.

	Level of importance						
Transport characteristics/ Level of importance	Unim-portant [%]	Slightly unimportant [%]	Less important [%]	Important [%]	Somewhat important [%]	Relatively important [%]	Most important [%]
1. Price	0	0	0	25	10	30	35
2. On-time performance/Adherence to the timetable	0	5	0	10	5	25	55
3. Quality/Reliability	0	0	5	10	20	45	25
4. Comfort and amenities	0	0	20	10	25	35	10
5. Speed of arrival (to the destination)	0	0	0	5	10	50	35
6. Punctuality/A regular schedule	0	5	0	0	20	50	25
7. Environmentalism/Ecology	5	5	20	5	15	35	15
8. Cleanliness of transport	0	5	5	0	35	40	15
9. Integrated payment system	0	5	0	5	25	40	25
10. Infrastructure of stops	0	5	0	5	35	30	25
11. Passengers' waiting time	0	0	5	0	35	50	10
12. Safety and security	0	0	0	5	10	45	40

Table 2 Transport characteristics evaluated by passengers

Definition of Inputs and Outputs

Based on the careful analysis of the obtained data, price, time, and carbon footprint variables were chosen as inputs, and the frequency of transport combinations was used as an output. The carbon footprint was calculated based on transport individuals, with specific transport modes [16] for each combination of DMUs; Time on its own was determined conventionally by covering a 5-kilometer by different types of vehicles and their average duration for each combination; For regional and city transport, the timeframe was calculated according to IDS-JMK. The average price has emerged based on a one-hour drive in one direction by different types of transport combinations; The final data doesn't include the cost of the vehicle itself or amortization expenses. The price for a private car was calculated according to the price of fuel. The cost of a taxi, a rented car, and a shared car was verified with a private company and is calculated for two people. Output is calculated as follows:

$$OUT1 = 1/4 \sum_{r1}^{5} v_r * f_r \qquad (2)$$

where v_r is weight for frequency *r (given by expression (1))* and f_r frequency (given in Table 1**).** Resulting output represents the normalized frequency of combination per passenger per day. The final result is given in Table 3**.** For example, for DMU1: $OUT1_{DMU1} = \frac{\sum_{r=1}^{5} v_r * f_r}{4} = \frac{(0.25*0)+\left(0.4*\frac{1}{30}\right)+\left(0.35*\frac{1}{4}\right)+\left(0*\frac{5}{7}\right)+(0*1)}{4} = 0.025$. f_{1DMU1} isn't considered in the calculation once assigned coefficient to f_1 has zero value.

DMUs	Transport combinations	In1	In2	In3	Out1
DMU1	Regional Transport (Train / Bus) & Brno City Public Transport (Bus / Tram / Trolleybus)	68	13	43	0.025
DMU2	Regional Transport (Train / Bus) & Walk (<1km / 15 min)	37	37	31	0.031
DMU3	Regional Transport (Train / Bus) & Bicycle	47	14	31	0.014
DMU4	Regional Transport (Train / Bus) & Shared Electric car / car / Taxi	73	7	498	0.008
DMU5	Own car & Brno City Public Transport (Bus / Tram / Trolleybus)	122	25	169	0.042
DMU6	Own car & Shared electric car /car / Taxi	127	7	623	0.010
DMU7	Own car & Bicycle	101	14	156	0.023
DMU8	Own car & Walk (<1km / 15 min) / own e-bike / e-Scooter / Hoover Boards etc.	108	22	189	0.029

Table 3 Original inputs and output of DMUs

In1 – Carbon footprint per passenger km (gCO2e/km); In2 – Time required to cover 2.5 km (min/km);

In3 – Costs required for a one-hour drive (CZK/km); Out 1 – Frequency of transport combination usage (comb/day);

3 Results and Discussion

Our study aims to discuss the potential of DEA method for solving the problem. In the literature we revealed mixed usage of both CCR and BCC, and various opinions on correctnes of VRS and CRS assumptions. Therefore, in this initial stage of research, we decided to apply consequently both methods and discuss their results and potential. By using the DEA method, we aim to examine the transport effectiveness in the South Moravian Region. The method's application is demonstrated for 8 DMUs, 3 inputs, and 1 output. DEA-SOLVER-LV8 (2014) was used for calculations, the results of which are presented in Table 4. A comparison between original and projected data displayed in a table indicates that each of the inefficient DMUs can reach efficiency if variables are changed accordingly. The combination of the highest efficiency score and the lowest ranking number indicated DMUs as efficient. The results themselves differ depending on the model due to the scale issue between the CRS and VRS frontiers.

	DMUs	Original Data						Projection			
		In1	In 2	In 3	Out 1	Efficiency Score	Rank	In1	In 2	In 3	Out 1
CCR – O Model	**DMU1**	68	13	43	0.025	1	1	68	13	43	0.025
	DMU2	37	37	31	0.031	1	1	37	37	31	0.031
	DMU3	47	14	31	0.014	0.69	7	47	14	30.99*	0.020*
	DMU4	73	7	498	0.008	0.55	8	36.62*	7	23.15*	0.014*
	DMU5	122	25	169	0.042	0.91	3	122	25	77.57*	0.046*
	DMU6	127	7	623	0.010	0.72	5	36.62*	7	23.15*	0.014*
	DMU7	101	14	156	0.023	0.84	4	73.23*	14	46.31*	0.027*
	DMU8	108	22	189	0.029	0.72	6	108	22	68.64*	0.041*
BCC – O Model	**DMU1**	68	13	43	0.025	1	1	68	13	43	0.025
	DMU2	37	37	31	0.031	1	1	37	37	31	0.031
	DMU3	47	14	31	0.014	1	1	47	14	31	0.014
	DMU4	73	7	498	0.008	1	1	73	7	498	0.008
	DMU5	122	25	169	0.042	1	1	122	25	169	0.042
	DMU6	127	7	623	0.010	1	1	127	7	623	0.010
	DMU7	101	14	156	0.023	0.85	7	72.5*	14	53.5*	0.027*
	DMU8	108	22	189	0.029	0.78	8	108	22	136.5*	0.038*

Table 4 CCR-O and BBC-O models' results and improvements

In the CCR-O model, only DMU 1 and DMU 2 are efficient, meaning they are on the frontier, while DMUs 3, 4, 5, 6, 7, and 8 are inefficient, meaning they are beneath of efficiency boundary. Among the worst performers outlined DMU 3 (2nd lowest output level) and DMU 4 (lowest output level). Based on the efficiency score, DMU 4 should increase its output by 45% in order to become efficient. The BCC-O model rated more DMUs as efficient; namely, there are a total of six efficient DMUs (1, 2, 3, 4, 5 and 6) meaning they are on the VRS efficient frontier. DMU 7 and DMU 8 are inefficient, rather underperforming compared to other combinations. However, their efficiency scores are relatively high (0.85 and 0.78) compared to CCR-O inefficient DMUs. The DMU 8 efficiency score is the lowest (78%), meaning that the DMU should augment its outputs by 22% to reach efficiency. According to the results it would be advisable to apply CCR model. For the efficient units also, super efficiency evaluation might be beneficial in the future research. Note that in future research the problem will be solved also using multiple criteria optimization to provide more robust results.

4 Conclusion

This paper represents preliminary research which applies Data Envelopment Analysis (output-oriented CCR and BCC models) to find an efficient transport design for the South Moravian Region. Accordingly, eight different public, regional, and private transport combinations were selected for estimation. Vital data was obtained based on a trial survey among residents of the region, which unequivocally pointed out that travel time, cost, and security are the main motives and demands for passengers. Considering modern challenges tend to involve the environmental friendliness of transportation processes, the carbon footprint variable was also used for the analysis. Models

with the same data provided two different assessments. However, both output-oriented models simultaneously revealed DMU 1 and DMU 2 combinations as efficient. Namely, a combination of "Regional Transport & Brno City Public Transport and a combination of "Regional Transport & Walk." The efficiency of DMU 1 can be affected by below-average input data and above average output data. To explain in more detail, the frequency of this transport combination is above average, emits lower carbon emissions, spends comparable less time and is quite budget-friendly. In case of DMU 2, the variables of input data are almost the same, while output is higher than for DMU 1. If we sum up, the CCR model gave more accurate results than the BCC model. We can say that for current case only CCR model would be more relevant and acceptable to use.

It is also essential to highlight the limitations of research. (i) The pilot survey was intended for a short audience; for future research, data will be collected for thousand passengers; (ii) In the study, one output variable was used from the passenger's perspective; further studies will extend variables as well as DMUs from various stakeholders' viewpoints, (iii) in the preliminary research basic DEA models were used. Future expanded studies will include more precise and complex models to achieve the finest results. Application of advance slack-based models (SBM) could be another possibility how to do analysis with more variables. It would be also beneficial to include some MCD method to confirm or improve the results of the DEA models.

Acknowledgements

The work was supported by the project No. IGA-PEF-DP-23-030 of the Internal Grant Agency, Faculty of Business and Economics, Mendel University, Brno.

References

[1] Bajec, P., Kontelj, M. & Groznik, A. (2020). Assessment of Logistics Platform Efficiency Using an Integrated Delphi Analytic Hierarchy Process-data Envelopment Analysis Approach: A Novel Methodological Approach Including a Case Study in Slovenia. *E & M Ekonomie A Management*, 23(3), 191-207. https://doi.org/10.15240/tul/001/2020-3-012.

[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. https://doi.org/10.1287/mnsc.30.9.1078.

[3] Charnes, A., Cooper, W.W. & Rhodes, E. (1978). Measuring the Inefficiency of Decision-Making Units. *European Journal of Operational Research*, 2, 429–444 ISSN 0377-2217. https://doi.org/10.1016/0377-2217(78)90138-8.

[4] Darold T. Barnum, et al. (2007). Comparing the Efficiency of Public Transportation Subunits Using Data Envelopment Analysis. *Journal of Public Transportation*, 10(2), 1-16. https://doi.org/10.5038/2375-0901.10.2.1.

[5] European Commission, (2011a). White Paper. Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system. COM (2011) 144 final, Brussels

[6] Fitzová, H., Matulová, M. & Tomeš, Z. (2018). Determinants of urban public transport efficiency: case study of the Czech Republic. *European Transport Research Review* 10, 42. https://doi.org/10.1186/s12544-018-0311-y.

[7] Fitzová, H. & Matulová, M. (2020) Comparison of urban public transport systems in the Czech Republic and Slovakia: Factors underpinning efficiency, *Research in Transportation Economics*, 81(C), https://doi.org/10.1016/j.retrec.2020.100824.

[8] Hajduk, S. (2018). Efficiency evaluation of urban transport using the DEA method. Equilibrium. *Quarterly Journal of Economics and Economic Policy*, 13(1), 141–157. https://doi.org/10.24136/eq.2018.008.

[9] Han, Z., Liao, L. & Wang, G. (2018). Research on performance evaluation method of Public Transit Routes based on BCC model. *Filomat*. 32, 1887-1896. https://doi.org/10.2298/FIL1805887H.

[10] Huguenin, J.-M. (2012). Data Envelopment Analysis (DEA): a pedagogical guide for decision makers in the public sector. Cahier de l'IDHEAP No. 276. *Lausanne: Swiss Graduate School of Public Administration* ISBN 978-2-940390-54-0

[11] Janová, J., Vavřina, J. & Hampel, D. (2012). DEA as a tool for bankruptcy assessment: the agribusiness case study. In: *Proceedings of the 30th International Conference Mathematical Methods in Economics*, 379-383.

[12] Janová, J., Hampel, D., Kadlec, J. & Vrška, T. (2022). Motivations behind the forest managers' decision making about mixed forests in the Czech Republic. *Forest Policy and Economics*. 144, 102841 https://doi.org/10.1016/j.forpol.2022.102841.

[13] Kaisar, E.I., Teegavarapu, R.S. & Gundersen, E. (2019). Data Envelopment Analysis Model for Assessment of Safety and Security of Intermodal Transportation Facilities. *Journal of Traffic and Transportation Engineering*, 7, 191-205. https://doi.org/10.17265/2328-2142/2019.05.001.

[14] Leal Filho, W., Ng, A.W., Sharifi, A. et al. (2023). Global tourism, climate change and energy sustainability: assessing carbon reduction mitigating measures from the aviation industry. *Sustainability Science*, 18, 983–996. https://doi.org/10.1007/s11625-022-01207-x.

[15] Li, Q., Bai, P. R., Chen, Y. & Wei X. (2020). Efficiency Evaluation of Bus Transport Operations Given Exogenous Environmental Factors, *Journal of Advanced Transportation*, 1-13 https://doi.org/10.1155/2020/8899782.

[16] Ritchie, H. (2020). "Which Form of Transport Has the Smallest Carbon Footprint?" [Online]. Available at: https://ourworldindata.org/travel-carbon-footprint [cited 2023-05-14].

[17] Staňková, M., Hampel, D. & Janová, J. (2022). Micro-Data Efficiency Evaluation of Forest Companies: The Case of Central Europe. *Croatian journal of forest engineering*, 43, 441-456. https://doi.org/10.5552/crojfe.2022.1541.

[18] Sujakhu, S. & Li, W. (2020). Public Transit Performance Evaluation Using Data Envelopment Analysis and Possibilities of Enhancement. *Journal of Transportation Technologies*, 10, 89-109. https://doi.org/10.4236/jtts.2020.102006.

Statistical Analysis of Determinants of Military Recruitment

Alžběta Heroschová[1], Marek Sedlačík[2], Kamila Hasilová[3], Jakub Odehnal[4]

Abstract. The conclusions of the NATO Summit in Wales in 2014 and the subsequent change in the security situation entail a requirement for an increase in the number of members of the Czech Army, which competes for human resources in the labour market. This paper focuses on the problem of identifying and analysing the determinants of military recruitment affected by the development of the country's economic environment, measured through economic growth, labour market equilibrium, or wages. Statistical methods including the stepwise regression were used to select variables influencing the military recruitment. The results of parametric and nonparametric regression analyses confirm a positive effect of the wages, and a negative effect of the labour market equilibrium. Moreover, the economic growth has no statistically significant effect on military recruitment.

Keywords: military recruitment, determinants, regression, correlation

JEL Classification: C49
AMS Classification: 62P20

1 Economic Determinants of Military Recruitment

At present, the security environment is going through major changes, which implies that NATO member countries are required to increase military spending in order to ensure their own defence. On that account, the increase in military spending aimed at fulfilling the political commitment to spend 2% of the country's gross domestic product on defence naturally leads to an increase in allocated financial resources (factors influencing the level of military expenditures are described in [7, 8]), but also to pressure to increase the number of personnel. The requirement to increase the number of members of the armed forces thus represents a factor influencing the demand in the labour market, especially the demand for specific professions requiring the necessary physical readiness, specific knowledge and willingness to serve in the armed forces. Finding an appropriate balance in the labour market is influenced by a number of economic and non-economic factors motivating individuals in their decision to join the armed forces. The economic factors that can directly motivate individuals include, above all, the price of labour, especially when compared with the development of the average price of labour in the labour market in the Czech Republic, or the minimum wage, which can negatively motivate the supply in the labour market if it is higher that the equilibrium value. At the same time, it is also possible to expect an indirect effect of economic development measured by the gross domestic product growth rate; especially in times of economic recession, and during periods of GDP decline and subsequent increase in unemployment rate, it is possible to expect individuals to be more willing to serve in the armed forces, since the military may be perceived as a stable employer in the labour market.

The analysis of economic variables (gross domestic product growth rate, unemployment rate, average wage, minimum wage) defined as possible determinants of the labour market specifically affecting labour supply; i.e., the willingness of individuals to serve in the armed forces, is the subject of a number of scientific publications, see [2, 6, 11]. Lescreva [6] analyses the impact of the unemployment rate on success of recruitment into the armed forces. The findings of the study reveal the fact that the period of economic recession manifested by high unemployment rates has a positive effect on the recruitment of new staff to the armed forces. By contrast, periods of economic growth, and therefore decreasing unemployment rates, have a negative impact on military recruitment. Furthermore, the author discusses measures which could positively influence the recruitment process, including investment in marketing activities, increasing the number of recruiters, and increasing recruitment allowances. A similar economic context of the labour market is described in a paper by Bäckström [2], which primarily analyses the

[1] University of Defence, Brno, Department of Resource Management, Kounicova 65, Brno, alzbeta.heroschova@unob.cz
[2] University of Defence, Brno, Department of Quantitative Methods, Kounicova 65, Brno, marek.sedlacik@unob.cz
[3] University of Defence, Brno, Department of Quantitative Methods, Kounicova 65, Brno, kamila.hasilova@unob.cz
[4] University of Defence, Brno, Department of Resource Management, Kounicova 65, Brno, jakub.odehnal@unob.cz

impact of the unemployment rate on the rate of applications for the Swedish army recruitment process, and the impact of the differences between civilian and military professions in the labour market, which he sees primarily in the price of labour and, therefore, in the respective remuneration for service in the army, and in the form of non-monetary benefits associated with military service. The Swedish society sees the non-monetary benefits of the military service mainly in the opportunities to get stationed abroad, in employment security, in satisfaction derived from patriotism, family traditions or, for example, desire for adventure that the military profession promises to candidates. A factor discouraging applicants from joining the armed forces may be the strenuous work, the risk of harm to health or loss of life, stress, and the demands resulting from commitment and loyalty to the military. The results of an analysis of regional data containing information from 21 Swedish regions collected between 2011 and 2015 reveal a positive effect of the unemployment rate on military recruitment in Sweden. Consequently, Swedish regions with high unemployment rates were identified as having a high number of applications to join the army, thus a higher interest in serving in the army compared to regions with lower unemployment rates. Similarly, Warner [11] analysed the impact of the economy on recruitment and retention in the United States military. He believes that in order to successfully recruit and retain personnel, it is important to be able to make the military more attractive than competing sectors in the labour market. In order to achieve this goal, the army must first and foremost focus on providing sufficient compensation, in terms of both monetary compensation and also payment in kind (e.g. health care). The study concludes that job applicants and active-duty soldiers respond to military pay levels and wage rates in the labour market, as well as to recruitment allowances and bonuses. The results of the study show that if the military pay increases by 10%, there is 6-11% increase in the supply of highly qualified applicants to the armed forces. The author claims that if the unemployment rate declines by 10%, the number of applicants for military service decreases by 2–4%. Selected socio-economic determinants influencing the process of recruitment of personnel into the armed forces will be analysed in the following part of the paper by means of a regression model, which allows us to identify possible links between individual variables.

2 Data and Methods

For the purpose of the analysis, variables characterising the number of newly recruited professional soldiers in the analysed period from 2004 to 2021 were used in the form of a variable expressing the total increments of all members of the armed forces, as well as increments in the age category of 18 to 30 years of age, and at the same time increments of the number of female professional soldiers. Figure 1 implies that during the analysed period, it is possible to observe a significant decline, especially in the period of the global economic crisis, the consequences of which significantly affected the government spending in the Czech Republic, hence the size of military spending, which was significantly reduced. This was also negatively reflected in the decline in new recruits (in both categories) to the armed forces.

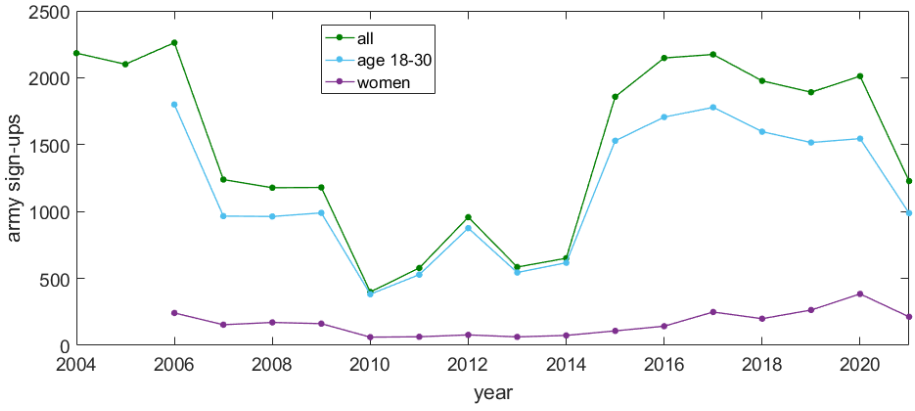


Figure 1 Increase of personnel of the armed forces, total number of new hires (green line), new hires in the 18-30 age group (blue line), and newly hired women (violet line).

As determinants of the recruitment process, variables characterizing the economic performance of the country measured by the evolution of the gross domestic product, the unemployment rate, the size of the minimum wage, and the ratio of the salary of a professional soldier to the average wage were selected. The model also includes an analysis of factors influencing the process of recruitment of women into the armed forces, where the relationship between the number of women in the army and the number of new female soldiers is analysed. Based on the defined determinants of the recruitment process, the following section will analyse the relationship between the

selected determinants and the newly recruited members of the armed forces (total numbers, age category 18 to 30 years, women). The analysis will focus mainly on the impact of economic cycles, and therefore, on the expected link between GDP and the number of new members of the armed forces, where we expect an increase in interest in serving in the armed forces, especially in a period of decreasing GDP; i.e., in a period of rising unemployment rate. Similarly, the link between the number of newly recruited members of the armed forces and the price of labour will be analysed. This is defined using a variable describing the ratio of the average salary of professional soldiers to the general average wage and the size of the minimum wage. Furthermore, the analysis is focused also on the influence of women in the military on the recruitment process itself.

In order to describe and study the dependency, the following regression model was applied: $Y_i = m(x_i) + \varepsilon_i$ for $i = 1, ..., n$. First of all, a linear regression model

$$m(x) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \tag{1}$$

was applied. In case of insignificance of the model or its coefficients (b_1, ..., b_k), it was reduced by stepwise regression. Then nonparametric regression model

$$m(x, h) = \sum_{i=1}^{n} \frac{K_h(x_i - x) y_i}{\sum_{i=1}^{n} K_h(x_i - x)}, \tag{2}$$

based on locally constant kernel estimates [3, 5] was employed. K_h is a kernel function and h is a smoothing parameter [10]. We applied the Gaussian kernel and selected the smoothing parameter according to an averaging procedure [1]. Subsequently, the calculations were supplemented with correlation analysis, for which we chose Pearson and Spearman correlation coefficients.

3 Results

The hypothesis for the analysis of the dependencies between the selected variables was that the main determinants influencing the recruitment process include GDP, the unemployment rate, the minimum wage, and the ratio of the soldier's salary to the average wage. This hypothesis was confirmed, but there is hidden collinearity between the variables. Using stepwise regression, we specified a submodel with determinants which play a role in the decision to join the army; i.e., the unemployment rate, and the wage ratio. The resulting model (1) has the form of

$$army\ sign\text{-}ups = -3442.7 - 98.642 \cdot unemployment\ rate + 48.272 \cdot ratio\ of\ wages$$

This model explains 54 % of the variability in the data ($R^2 = 0.54$), and is statistically significant at the 5% significance level ($p = 0.0012$). Figure 2 shows the graphical representation of the model.

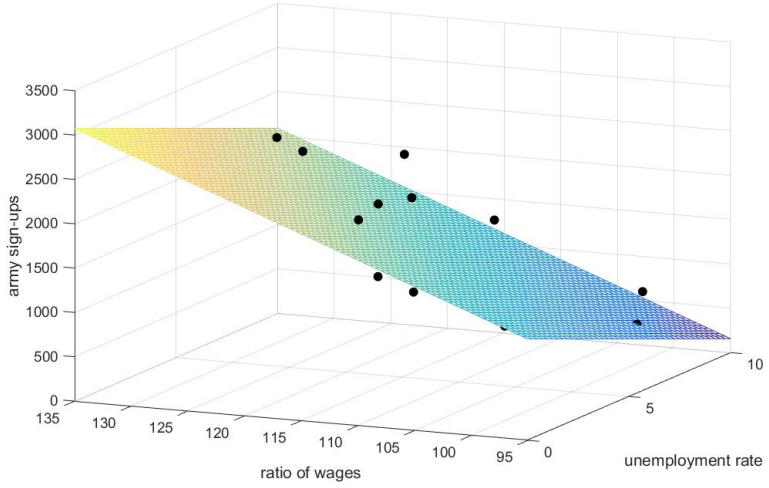


Figure 2 Linear model of the unemployment rate and the ratio of wages influencing the number of army sign-ups

The results of the model show that when the unemployment rate increases, the number of new members of the armed forces decreases; on the other hand, a higher ratio of the soldier's salary to the average salary attracts more people into the armed forces.

Further analysis focused on the group of people between 18 and 30 years of age. In this group, a significant factor influencing enlisting is the unemployment rate of this age group. Here we see a similar conclusion to the general model; i.e., that higher unemployment rates have a negative effect on entering the army. The corresponding equation of the linear model (1) is: *young army sign-ups* = 2287.6 −139.59·*young unemployment rate*, which is statistically significant (p = 0.0004), and explains 74% of data variability, as shown in Figure 3. Correlation analysis also confirmed the correlation between the unemployment rate and joining the army for this age group; specifically, Pearson $\rho = -0.8752$, and Spearman $\rho_S = -0.7182$, both statistically significant at the 5% level, p-values are 0.0004 and 0.0168, respectively.

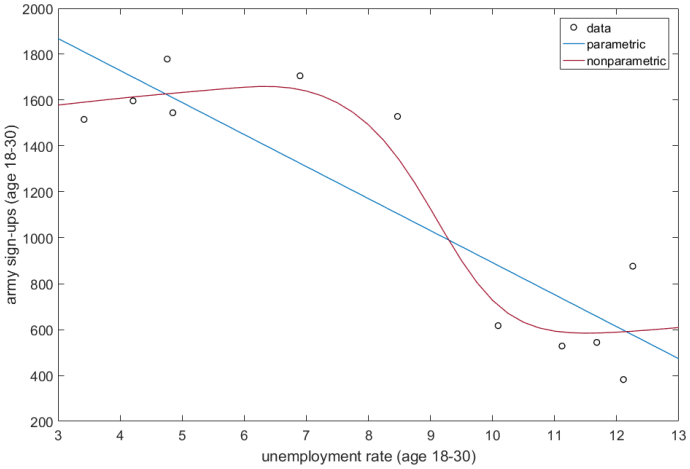


Figure 3 Model of the army sign-ups and the unemployment rate in the group of 18-30 year-olds (linear model is in blue, nonparametric model is in red)

Figure 3 also shows the nonparametric model (2) describing the relationship between the number of new recruits and the unemployment rate for the age group of 18 to 30-year-olds. It also shows an inversely proportional relationship between the number of newly recruited professional soldiers and the unemployment rate in this age category, while the effect of the economic crisis is also evident (compared to the linear model).

Similarly, the effect of gender on joining the military was analysed; in particular, we focused on whether the number of female professional soldiers has an effect on more women joining military. After excluding outliers, we obtained a linear model (1) with the equation *female army sign-ups* = −1054.2 +0.39482·*women in the army*, which is captured in Figure 4. Again, this is a statistically significant model (p = 0.0048) that explains almost half of the variability in the data (R^2 = 0.46). The correlation coefficients also indicate a correlation between these variables; to be specific, Pearson ρ = 0.7055 and Spearman ρ_S = 0.6923, with corresponding p-values of 0.0048 and 0.0079, respectively.

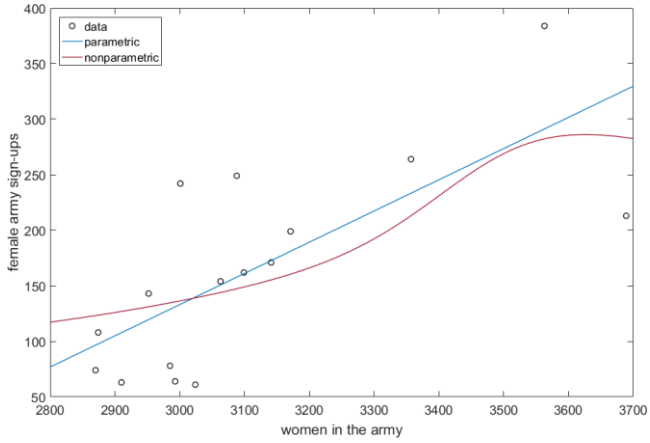


Figure 4 Model of army sign-ups among women (linear model in blue, nonparametric model in red)

Both parametric and nonparametric models show that a higher number of women in the military attracts more female recruits. There is, however, a component of randomness present that appears to be influenced by another (unmeasured) variable, but there is still a significant effect of this determinant (i.e., the number of women in the military) on joining the military.

The results of the model confirm the conclusions of Bäckström [2], Warner [11] about the possible positive effect of wages on the labour market decision of individuals to join the armed forces. At the same time, they do not confirm the expected negative relationship between a country's economic development and the number of new recruits to the armed forces (for more details see [4]), which was described, for example, in [6] using the example of the USA.

4 Conclusions

The Armed Forces of the Czech Republic are an important element of the country's security. This is especially true in today's uncertain times, when not only the country, but also all NATO and EU member states, and other European countries are directly or indirectly threatened by a great deal of both state and non-state actors [9]. The readiness of the armed forces to respond to possible threats of military or non-military nature is necessary, and they must operate in the required numbers and quality. A possible increase in the number of members of the armed forces thus creates additional demand in the labour market, which is naturally influenced by individual determinants of the labour market. Selected socio-economic variables characterising the economic development of the country and its impact on the recruitment process, the price of labour and its impact on the recruitment process, the unemployment rate and the gender structure of the armed forces were analysed as determinants of the recruitment process. The results of the linear regression model thus reveal that the process of recruitment to the armed forces in the analysed period between 2004 and 2021 was positively influenced primarily by the price of labour, where in the case of an increase in the ratio between the price of a professional soldier's labour and the average wage, there is also an increase in interest in serving in the armed forces, which is manifested in the number of newly recruited soldiers. At the same time, the positive effect of the number of women in the army on new recruits was confirmed. Furthermore, the analysis did not prove the expected impact of economic development on the recruitment process, since the armed forces did not act as an employer with a demonstrable increase in people interested in serving in the armed forces during a period of economic downturn and rising unemployment. The findings thus clearly demonstrate wages as one of the main factors influencing the attractiveness of the armed forces as an employer in the labour market.

Acknowledgements

The paper was supported by institutional funding aimed at the development of the research organization; i.e., Faculty of Military Leadership of the University of Defence.

References

[1] Baszczyńska, A. K. (2017). One value of smoothing parameter vs interval of smoothing parameter values in kernel density estimation. *Acta Universitatis Lodziensis. Folia Oeconomica*, 6: 73–86.

[2] Bäckström, P. (2018). Are Economic Upturns Bad for Military Recruitment? A Study on Swedish Regional Data 2011–2015. *Defence and Peace Economics*. Routledge, 2018, 30(7), 813–829.

[3] Hasilová, K. & Vališ, D. (2022). Composite laminates reliability assessment using diffusion process backed up by perspective forms of non-parametric kernel estimators. *Engineering Failure Analysis*, 138: 106326.

[4] Holcner, V., Davidová, M., Neubauer, J., Kubínyi, L. & Flachbart, A. (2021). Military Recruitment and Czech Labour Market. *Prague Economic Papers*, 2021(4), 489–505.

[5] Horová, I., Koláček, J. & Zelinka, J. (2012). *Kernel smoothing in Matlab: Theory and practice of kernel smoothing*. World Scientific, Singapore.

[6] Lescreve, J. F. (2022). *Recruiting for the Military when the Economy is booming*. 36th International Applied Military Psychology Symposium Split, Croatia. [Online]. Available at: http://www.iamps.org/IAMPS_2000_Lescreve_Recruiting_for_the_Military_.pdf [cited 2022-12-04].

[7] Neubauer, J. & Odehnal, J. (2018). Security and Economic Determinants of the Demand for Czech Military Expenditure: ARDL Approach. In: *International Conference of Numerical Analysis and Applied Mathematics (ICNAAM 2017) AIP Conf. Proc.* 1978. Melville: AIP Publishing.

[8] Odehnal, J. & Sedlačík, M. (2015). The demand for military spending in NATO member countries. *AIP Conference Proceedings*. Melville, New York: American Institute of Physics.

[9] Odehnal, J. & Sedlačík, M. (2018) Political stability as determinant of terrorist attacks in developed and developing countries: An empirical multivariate classification analysis. In: *AIP Conference Proceedings.* Greece: American Institute of Physics.
[10] Wand, M. P. & Jones, M. C. (1995). *Kernel smoothing*. Chapman and Hall, London.
[11] Warner, T. J. (2012). *The Effect of the Civilian Economy on Recruiting and Retention. The Eleventh Quadrennial Review of Military Compensation*. Washington, D.C.: Department of Defense: Office of the Under Secretary od Defense for Personnel and Readiness, 2012, 6/2012, 71–91. [Online]. Available at: https://militarypay.defense.gov/Portals/3/Documents/Reports/SR05_Chapter_2.pdf [cited 2022-11-28].

Designing an Efficient Transportation System under a Constrained Budget and Cost Uncertainty

Robert Hlavatý[1], Helena Brožová[2]

Abstract. We revisit a particular case of the traditional transportation problem of Hitchcock and Koopmans with multiple objective functions. We assume that the maintenance costs of each supplier influence the capacities on the suppliers' side. Further, we assume that the maintenance costs are subject to uncertainty and can vary within a predefined range. We seek to design an optimal transportation system with respect to all objectives while arranging the capacities of the suppliers considering a given budget and uncertainty in the maintenance costs. We propose an algorithm based on the De Novo optimization concept, which yields an efficient constellation of the suppliers' capacities under the constrained budget. To capture the uncertainty within the budget, we employ the concept of gamma-robust optimization. We illustrate the algorithm on a simple artificial case. It turns out that the proposed approach allows for efficient budget use and offers robust solutions that face the financial uncertainty of the decision-maker.

Keywords: De Novo Optimization, Multi-objective Linear Programming, Robustness, Transportation problem, Uncertainty

JEL Classification: C61, R41
AMS Classification: 90C17, 90C29

1 Introduction

In this paper, we continue our former investigation of the traditional transportation problem (TTP) initially developed by Hitchcock [5] and Koopmans [8]. The traditional transportation problem in its original form describes a task of allocating a homogeneous commodity on the routes between a set of suppliers to a set of destinations while minimizing the cost for such an allocation. We have recently provided a new perspective on the traditional transportation problem with multiple objectives. Instead of searching for an efficient multi-objective solution under the given constraints, we [7] focused on the efficiency of the transportation system design itself. In our [7] recent contribution, we have investigated the efficiency of resource allocation on the supplier's side when the unit-wise maintenance costs are set for each supplier. Our results showed that in the case of a multi-objective transportation problem (MOTP), the supplier's capacities can be rearranged so that better objective values can be achieved under the constrained original budget. Therefore, we proposed a way to efficient transportation system design that achieves better performance towards all objectives that are subject to optimization (in comparison with the original system design). The essence of our approach was inspired by the concept called *De Novo* optimization, originally presented by Zelený [11]. In the current contribution, we consider that the maintenance costs of suppliers are subject to uncertainty. A methodology called *gamma-robustness* by Bertsimas and Sim [2] can be used for this purpose. We have earlier investigated the TTP with the interval-based cost coefficients and presented a way to utilize the *gamma-robustness* to obtain a robust (i.e., guaranteed) solution under a given degree of uncertainty [6].

The current paper aims to show how to design an efficient MOTP with respect to a constrained budget by efficiently rearranging the supplier's capacities while the unit maintenance costs are not fixed, and we allow cost variations within the predefined intervals. As a result of our investigation, we present an efficient transportation system design that respects cost deviations and provides a solution protected against uncertainty. The uncertainty in the efficient system design has already been studied specifically for the *De Novo* methodology by Liu and Shi [9] and Chen and Hsieh [4], but both papers are based on using fuzzy arithmetic to capture the uncertainty in the model input. We further present an approach where the uncertainty is measured by an algebraic function instead. We begin with a short overview of the MOTP, the *De Novo* methodology, and *gamma-robust* optimization. We implement both methodologies in the MOTP and provide a brief artificial example to illustrate its potential at the end of this paper.

[1] CZU Prague, Department of Systems Engineering, Kamycka 129, 16500 Prague – Suchdol, hlavaty@pef.czu.cz
[2] CZU Prague, Department of Systems Engineering, Kamycka 129, 16500 Prague – Suchdol, brozova@pef.czu.cz

2 Methods

In the methodology part, we provide a generic MOTP formulation to clarify the further notation, and afterwards, we explain the essentials of the *De Novo* approach and *gamma-robustness*.

2.1 Multi-objective transportation problem

For the purposes of our research, we need a linear programming formulation of MOTP. The problem has m suppliers who provide a homogeneous commodity to n destinations. Each supplier $i = 1,..,m$ provides a supply a_i, each destination $j = 1,..,n$ has a demand b_j and we consider that the problem is balanced on both sides such that $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$. In general, the balance assumption is not crucial to our approach and can be simply overcome by adding a dummy component to the problem. However, to avoid the overload of notation, we consider the balanced version furthermore. We assume p objectives that are distinguished as two index sets K^{Max}, K^{Min} according to their form. The multi-objective linear optimization model of MOTP is formulated as follows:

$$\begin{aligned} &\max \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Max} \\ &\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Min} \\ \text{s.t.}\quad & \\ &\sum_{j=1}^{n} x_{ij} = a_i\,, i = 1,\dots,m \\ &\sum_{i=1}^{m} x_{ij} = b_j\,, j = 1,\dots,n \\ &x_{ij} \geq 0, \forall i,j \end{aligned} \tag{1}$$

We seek the efficient solution $z^T = (z^{Max}, z^{Min}) = (z_1, \dots, z_p)$ that respects the objective costs c_{ijk} and corresponds to the transported amount of the commodity which is for each route given by the value of x_{ij}.

2.2 *De Novo* optimization approach

The general framework of the *De Novo* optimization approach is used for an efficient system design, where the system is described as a linear optimization problem. The original idea was first introduced by Zelený [11], but we use a more general form of the problem described by Brožová and Vlach [3]. We assume the following generic multi-objective optimization problem (MOOP):

$$\begin{aligned} &\max(C^{Max}x) \\ &\min(C^{Min}x) \\ \text{s.t.}\quad & \\ &Ax \leq b \\ &x \geq 0 \end{aligned} \tag{2}$$

We seek an efficient solution for p objectives with coefficients $(C^{Max}|C^{Min})^T \in \mathbb{R}^{p\times n}$ and we assume that the polyhedral set $Ax \leq b, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ is bounded and non-empty, and the expression shelters all possible relations to the right-hand side $(\leq, \geq, =)$. The principal idea of the *De Novo* concept is finding an optimal constellation of components b_i of b under a given maintenance budget B with the technical coefficients A. This means that the vector b does not attain predefined values and instead becomes a variable which we denote as $y \in \mathbb{R}^m, y \coloneqq b$. At the same time, we introduce a vector $q \in \mathbb{R}^m$ whose components q_i express unit maintenance cost of respective b_i. Given the constrained budget B, the feasible distribution of maintenance costs is $q^T y \leq B$. Considering this inequality, we obtain a new MOOP formulation with the unknown components of b that are optimized with respect to their costs q and an available budget B:

$$\begin{aligned} &\max(C^{Max}x) \\ &\min(C^{Min}x) \\ \text{s.t.}\quad & \\ &Ax - y \leq 0 \\ &q^T y \leq B \\ &x \geq 0 \end{aligned} \tag{3}$$

Problem (3) can be solved by any MOO approach; however, different methods would yield different values of y (and thus b). Therefore, the description of the specific MOO solution approach follows, which is based on the *De Novo* framework principles. First, we solve the problem (3) as a single-objective problem p times and obtain the vector of objective function values of the ideal solutions $z^{*T} = \left(z^{*Max}, z^{*Min}\right) = \left(z_1^*, \dots, z_p^*\right)$. Knowing z^*, the ideal constellation of components would be the one minimizing the necessary budget needed for achieving the ideal values in all objectives. This is achieved by solving the following problem, as proposed by Zhuang and Hocine [12]:

$$\begin{aligned} &\min(q^T y) \\ \text{s.t.} \\ &Ax - y = 0 \\ &C^{Max}x \geq z^{*Max} \\ &C^{Min}x \leq z^{*Min} \\ &x, y \geq 0 \end{aligned} \tag{4}$$

If there exists a feasible realization of the problem (4), then it has an optimal solution (x^*, y^*) with the minimum budget $B^* = q^T y^*$. However, if the solution of problem (4) is infeasible, then it is necessary to relax the 2nd and 3rd constraint, i.e., to set such a vector of objective values $\hat{z} = (\hat{z}^{Max}, \hat{z}^{Min})$ for which the problem (4) is solvable. In order to obtain $\hat{z}$, we take advantage of the STEM method. The method is primarily used for solving MOOP problems by minimizing the distance between z^* and $\hat{z}$. The method was proposed by Benayoun et al. [1], and for our purposes, we only apply its one iteration to obtain any feasible $\hat{z}$. Apart from the original source, we describe the details of the STEM method for this particular problem in [7] and skip the detailed description here. Once the feasible $\hat{z}$ is obtained, a new equivalent of problem (4) is built:

$$\begin{aligned} &\min(q^T y) \\ \text{s.t.} \\ &Ax - y = 0 \\ &C^{Max}x \geq \hat{z}^{Max} \\ &C^{Min}x \leq \hat{z}^{Min} \\ &x, y \geq 0 \end{aligned} \tag{5}$$

By solving the problem (5), we obtain the optimal solution $(\hat{x}, \hat{y})$ that relates to a budget $\hat{B} = q^T \hat{y}$. If the budget $\hat{B}$ corresponds with the original budget B, then $\hat{x}$ is the efficient solution to the original problem (2) with the efficient objective values $C\hat{x}$ and optimal design of the right-hand sides $\hat{y}$. In case that $\hat{B} \neq B$, it is possible to obtain the corresponding solution through the linear transformation of the solution vectors. The transformation was proposed by Shi [10] and uses the *optimum-path ratio* $r = \frac{B}{\hat{B}}$ to obtain the corresponding efficient solution $r\hat{x}$ with the efficient objective values $rC\hat{x}$ and optimal design of the right-hand sides $r\hat{y}$.

2.3 Gamma-robustness

We use this specific concept of robust optimization developed by Bertsimas and Sim [2] called Gamma-robustness to capture the uncertainty in the optimization problems in this paper. The methodology itself is quite extensive, and due to lack of space, we only present its basics without proofs and refer the reader to the original authors or our former paper dedicated to robust transportation [6] for a detailed explanation.

The core of the concept is building a *robust counterpart* to a given optimization problem. The *robust counterpart* is an optimization problem on its own and finds a guaranteed optimal solution that considers possible deviations in the model input. We assume the following single-objective optimization problem (6) and its specific form of a *robust counterpart* (7):

$$\begin{aligned} &\max c^T x \\ \text{s.t.} \\ &\sum_{j=1}^{n} \left(a_{ij} + \delta_{ij}^a\right) x_j \leq b_i \,, i = 1, \dots, m \\ &x_j \geq 0 \,, j = 1, \dots, n \end{aligned} \tag{6}$$

$$\begin{aligned} &\max c^T x \\ \text{s.t.} \\ &\sum_{j=1}^{n} a_{ij} x_j + \Gamma_i w_i + \sum_{j \in \mathcal{U}_i} p_{ij} \leq b_i \,, i = 1, \dots, m \\ &w_i + p_{ij} \geq \delta_{ij}^a x_j \,, i = 1, \dots, m \,, \forall j \in \mathcal{U}_i \\ &w_i \geq 0 \,, i = 1, \dots, m \\ &p_{ij} \geq 0 \,, i = 1, \dots, m \,, \forall j \in \mathcal{U}_i \\ &x_j \geq 0 \,, j = 1, \dots, n \end{aligned} \tag{7}$$

Problem (7) is a linear optimization problem while it is considered that its technical coefficients a_{ij} may or may not deviate by δ_{ij}^{a}. It seeks the optimal solution with the guaranteed objective value provided that there is an expectation that Γ_i unspecified coefficients a_{ij} in i-th constraint will deviate from its original (expected) value by δ_{ij}^{a}. In addition, w_i and p_{ij} are auxiliary variables and $\mathcal{U}_i$ denotes a set of indices j of those a_{ij} for which the deviations are considered. The existence of the *robust counterpart* is based on the primal-dual properties of the problem (6). Note that we only consider the deviations in the left-hand side coefficients because we do not need any other for the purposes of our paper.

3 Results

The main idea of this contribution is to develop the optimal design of the transportation system with multiple objectives under the *De Novo* framework while considering the uncertainty in the unit maintenance costs of suppliers within the transportation system. We combine the three methodologies described in the previous chapter to build the mathematical models that lead to the optimal transportation system design. We apply the *De Novo* framework described in (2.2) to the MOTP (1) as a sequence of the following steps. First, we start with the reformulated MOTP (8), which considers the uncertainty in the capacities on the suppliers' side. This means that parameters a_i are replaced by the variables $y_i, \forall i$ and these variables must now fit in the given budget B considering the uncertain costs q_i that can deviate at most by δ_i^q. Note that the value δ_i^q must be set by a decision-maker. Next, we build the model (9), which is the *robust counterpart* to (8):

$$\begin{aligned}
&\max \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Max}\\
&\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Min}\\
\text{s.t.}\quad &\\
&\sum_{j=1}^{n} x_{ij} = y_i\,, i = 1, \dots, m\\
&\sum_{i=1}^{m} x_{ij} = b_j\,, j = 1, \dots, n\\
&\left(q_i^T + \delta_i^q\right) y_i \le B\\
&x_{ij} \ge 0, y_i \ge 0, \forall i, j
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\max \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Max}\\
&\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk}x_{ij}\,, k \in K^{Min}\\
\text{s.t.}\quad &\\
&\sum_{j=1}^{n} x_{ij} = y_i\,, i = 1, \dots, m\\
&\sum_{i=1}^{m} x_{ij} = b_j\,, j = 1, \dots, n\\
&\sum_{i=1}^{m} q_i^T y_i + \sum_{i \in \mathcal{U}_i} p_i + \Gamma w \le B\\
&w + p_i \ge \delta_i^q y_i\,, \forall i \in \mathcal{U}_i\\
&p_i \ge 0, \forall i \in \mathcal{U}_i\\
&w \ge 0, x_{ij} \ge 0, y_i \ge 0, \forall i, j
\end{aligned} \tag{9}$$

Note that also $\Gamma, 0 \le \Gamma \le$ m must be set by the decision maker, and it denotes how many values of q_i are expected to deviate from their expected value by at most δ_i^q. In the next stage, we solve the problem (9) for all single-objective functions $k = 1, \dots, p$ and obtain the vector of ideal solutions $z^{*T} = \left(z^{*Max}, z^{*Min}\right) = \left(z_1^*, \dots, z_p^*\right)$. We follow the *De Novo* framework (formula 4) and formulate the new problem (11) that minimizes the necessary budget while achieving the ideal values of objectives z^*. Note that in problem (11), the costs q_i in the objective function are still subject to uncertainty, and it needs to be dealt with again using the *robust counterpart* approach. In this case, it is necessary to turn the objective into a constraint so that the gamma-robustness approach is applicable. We perform the substitution of the objective function and change it into a constraint that is bounded by the substitution term:

$$\begin{aligned}
&\min(\Phi)\\
\text{s.t.}\quad &\\
&q_i^T y_i \le \Phi
\end{aligned} \tag{10}$$

After the substitution has been done, it is finally possible to formulate the last stage of the *De Novo* sequence of optimization problems which is described as (12). Problem (12) is only a reformulation of (11) where q is subject to uncertainty and was changed in the constraint form to allow the creation of the *robust counterpart*.

$$\min \sum_{i=1}^{m} q_i^T y_i$$

s.t.

$$\sum_{j=1}^{n} x_{ij} = y_i \,, i = 1, \dots, m$$
$$\sum_{i=1}^{m} x_{ij} = b_j \,, j = 1, \dots, n \qquad (11)$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk} x_{ij} \geq z^{*Max}, k \in K^{Max}$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk} x_{ij} \leq z^{*Min}, k \in K^{Min}$$
$$x_{ij} \geq 0, y_i \geq 0, \forall i, j$$

$$\min(\Phi)$$

s.t.

$$\sum_{i=1}^{m} q_i^T y_i + \sum_{i \in \mathcal{U}_i} p_i + \Gamma w \leq \Phi$$
$$w + p_i \geq \delta_i^q y_i \,, \forall i \in \mathcal{U}_i$$
$$\sum_{j=1}^{n} x_{ij} = y_i \,, i = 1, \dots, m$$
$$\sum_{i=1}^{m} x_{ij} = b_j \,, j = 1, \dots, n \qquad (12)$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk} x_{ij} \geq z^{*Max}, k \in K^{Max}$$
$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ijk} x_{ij} \leq z^{*Min}, k \in K^{Min}$$
$$p_i \geq 0, \forall i \in \mathcal{U}_i$$
$$z \geq 0, x_{ij} \geq 0, y_i \geq 0, \forall i, j$$

Problem (12) finds the optimal transportation system design under uncertain costs. In case the solution of (12) is not feasible, then it is necessary to employ the STEM method and then solve it again while replacing ideal objectives z^* by achievable objectives $\hat{z}$.

3.1 Example

We illustrate our approach on a small artificial example with unspecified goods to be transported between three suppliers and three destinations. We assume the balanced problem with two objectives $z_1 = c_1^T x$ that minimizes the total distance and $z_2 = c_2^T x$ that maximizes the prices for which the goods are sold in the different destinations. All units used in the example are dimensionless. We present our problem in the classic transportation table rather than the mathematical model to make it more comprehensive for a reader. The situation is depicted in Figure 1, and the objective coefficients $c_1(c_2)$ are in the upper-right corners.

	Destination 1	Destination 2	Destination 3	Supply
Supplier 1	5(6)	4(12)	1(7)	9
Supplier 2	3(8)	2(14)	6(9)	16
Supplier 3	7(10)	5(13)	8(11)	18
Demand	10	15	12	

Figure 1 The illustrative example in the transportation table

In the original situation, the supplier capacities are $b^T = (9, 16, 18)$, and we assume that the unit maintenance costs for each supplier are $q^T = (11, 17, 21)$. Next, we assume that b is not fixed, replace it with a variable y and seek the optimal constellation of y under costs q. Moreover, we assume that the costs are not fixed either and can vary by the predefined deviations $\delta^{qT} = (5.5, 8.5, 10.5)$. Here, we set these deviations such that $\delta_i^q = 0.5 q_i$ but the decision generally depends on the decision-maker. We assume that our budget is given by the expected costs and original capacity distribution: $B = q^T b = 749$. Next, we apply the methodology described in Chapter 2 and repeat the calculation for the different levels of $\Gamma = \{0,1,2,3\}$ respectively. The value of Γ used in the individual computations shows how many costs of q are expected to deviate by δ^q while it is not said which of them it would be. $\Gamma = 0$ means that no cost would deviate from its expected value, and $\Gamma = 3$ means that all costs would. These two scenarios are trivial and could be computed without using the *gamma-robustness* concept.

We obtain the solutions under the original budget B depicted in Figure 2 show the guaranteed objectives with the optimal distribution of the suppliers' capacities for each case. One can observe that with each degree of uncertainty, the guaranteed objectives differ, as well as the overall distribution of goods and suppliers' capacities. In the case of $\Gamma = 0$ or $\Gamma = 3$, we do not speak of uncertainty because either nothing or everything would change, and there is only one scenario for such a situation. On the other hand, for $\Gamma = 1$ or $\Gamma = 2$, we can see the optimal arrangement, which guarantees there will be no worse achievement than this. The total transported amount decreases with

the increasing uncertainty. For $\Gamma = 0$ or $\Gamma = 3$ we can see the highest transported amount, as the optimal solution under certainty is better organized in the suppliers' capacities under given budget B.

Γ=0	Destination 1	Destination 2	Destination 3	Supply
Supplier 1	5(6)	4(12)	1(7) 15.35	15.35
Supplier 2	3(8) 13.33	2(14) 20	6(9)	33.33
Supplier 3	7(10)	5(13)	8(11) 0.64	0.64
Demand	13.33	20	15.99	49.32
z_1=100.52		z_2=501.16		

Γ=1	Destination 1	Destination 2	Destination 3	Supply
Supplier 1	5(6)	4(12)	1(7) 10.98	10.98
Supplier 2	3(8) 9.65	2(14) 14.48	6(9)	24.13
Supplier 3	7(10)	5(13)	8(11) 0.61	0.61
Demand	9.65	14.48	11.59	35.72
z_1=73.77		z_2=363.51		

Γ=2	Destination 1	Destination 2	Destination 3	Supply
Supplier 1	5(6)	4(12) 0.69	1(7) 11.11	11.8
Supplier 2	3(8) 5.15	2(14) 13.2	6(9)	18.35
Supplier 3	7(10) 4.11	5(13)	8(11)	4.11
Demand	9.26	13.89	11.11	34.26
z_1=84.47		z_2=353.16		

Γ=3	Destination 1	Destination 2	Destination 3	Supply
Supplier 1	5(6)	4(12) 7.32	1(7) 11.62	18.94
Supplier 2	3(8) 8.69	2(14) 7.2	6(9)	15.89
Supplier 3	7(10) 0.99	5(13)	8(11)	0.99
Demand	9.68	14.52	11.62	35.82
z_1=88.32		z_2=349.45		

Figure 2 The robust solutions for the different levels of Γ

4 Conclusion

The example shows the possibilities of combining two optimization approaches enabling the optimization of the system structure. Therefore, it is also important to analyze the results with different degrees of uncertainty. We further aim to perform computational experiments with large-scale transportation cases to show how the level of uncertainty measured by Γ influences the costs of building such robust scenarios that protect the decision-maker against uncertainty.

Acknowledgements

This paper was supported by the Czech Science Foundation (GAČR) project No. P403-22-11117S.

References

[1] Benayoun, R., Montgolfier, J., Tergny, J. & Laritchev, O. (1971). Linear programming with multiple objective functions: Step method (STEM). *Mathematical Programming*, *1(1)*, 366-375.

[2] Bertsimas, D. & Sim, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming, 98*, 49-71.

[3] Brožová, H. & Vlach, M. (2018). Remarks on De Novo approach to multiple criteria optimization. In *36th International Conference on Mathematical Methods in Economics (MME)*, *Jindřichův Hradec*, 618-623.

[4] Chen, Y.-W. & Hsieh, H.-E. (2006). Fuzzy multi-stage De-Novo programming problem. *Applied Mathematics and Computation, 181*, 1139-1147.

[5] Hitchcock, F.L. (1941). The distribution of a product from several sources to numerous locations. *Journal of Mathematics and Physics*, *20,* 224-230.

[6] Hlavatý, R. & Brožová, H. (2017). Robust optimization approach in transportation problem. In *Proceedings of 35th International Conference Mathematical Methods in Economics (MME), September 13-15,* 225-230.

[7] Hlavatý, R. & Brožová, H. (2023). Metaoptimization approach to designing optimal transportation system with multiple criteria. In *Proceedings of the 15th International Conference on Strategic Management and its Support by Inńformation Systems, Ostrava*. (in print).

[8] Koopmans, T.C. (1947). Optimum utilization of the transportation system. In *Proceedings of the International Statistical Conference*, Washington, DC.

[9] Liu, Y.-H. & Shi, Y. (1994). A fuzzy programming approach for solving a multiple criteria and multiple constraint level linear programming problem. *Fuzzy Sets and Systems, 65*, 117-124.

[10] Shi, Y. (1995). Studies on optimum-path ratios in multicriteria de novo programming problems. *Computers & Mathematics with Applications, 29*, 43-50.

[11] Zelený, M. (1986). Optimal system design with multiple criteria: De Novo programming approach. *Engineering Costs and Production Economics, 10*, 89-94.

[12] Zhuang, Z.-Y. & Hocine, A. (2018). Meta goal programing approach for solving multicriteria de Novo programing problem. *European Journal of Operational Research, 265*, 228-238.

A User Recommendation System Based on Graph Neural Network and Contextual Behavior

Jiri Homan[1], Ladislav Beranek[2], Radim Remes[3]

Abstract. Today, recommendation systems are an integral part of e-commerce services on the Internet. In connection with their development, neural networks have become the most used approach to recommender systems. In our post, we will demonstrate the use of graph neural networks to create a recommender system. E-commerce systems can be modeled using a bipartite interaction graph. There are two essential parts to this chart, users and items. In our model, context is added to them and integrated into the mentioned parts of the bipartite graph using the theory of hypothetical functions. Different elements of a bipartite graph can interact using edges. Therefore, modeling the interaction of elements can be transformed into modeling the interaction of nodes on the corresponding graph. We implemented a recommender system model in Python and used relevant libraries, which we tested on standard datasets. These experiments showed the good ability of our model for recommendations. We used the root mean square error (RMSE) and mean absolute error (MAE) indicators.

Keywords: E-commerce, Recommender Systems, Graph Neural Networks, Belief Function Theory

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Recommender systems are essential for most online service platforms, such as e-commerce, social networking websites, and other online media. A recommendation system is essentially a kind of filtering system in which the goal is to present personalized information to users. Such a system improves user comfort. Its purpose is to suggest items to users that might interest them.

Recommender system algorithms can be divided into three groups:

- Collaborative filtering algorithms that offer items to the user based on historical interactions. These can be ratings or feedback as numerical ratings or text. These algorithms extract the mutual information between the user and the item to make a recommendation. They are mainly used for video, audio, and text data. The principle is the similarity of user preferences.
- Hybrid recommendation system algorithms can integrate in multiple ways when offering recommendations [22, 23]. These are user preferences, item characteristics, and other appropriate information representing the correlation relationship between users and items [19], [8]. Various products can be offered based on this information.
- The matrix factorization (MF) or singular value decomposition (SVD) method can find out the preferences of each user concerning individual items by optimizing the user-item matrix [12, 13]. However, finding the optimal values of such a matrix is time-consuming compared to previous item-based and user-based filtering algorithms.

Recently, neural networks have been used to solve collaborative filtering problems. Neural network-based models are proposed for solving problems such as complex user behavior or data entry [5, 10]. Examples can be neural collaborative filtering (NCF) or deep learning factorization (DeepFM) [10]. Graph neural networks appear as the current direction for solving recommender systems. They enable the manipulation of structural data and the exploration of structural information. GNN-based approaches have become the new state-of-the-art approaches in recommender systems.

[1] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, homan@ef.jcu.cz
[2] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, beranek@ef.jcu.cz.
[3] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, inrem@ef.jcu.cz.

This paper focuses on developing a collaborative filtering method that combines a graph neural (convolutional) network (GCN) model with reasoning based on Dempster-Shafer's assumption function theory for a multicriteria recommender system. The GCN model integrates a matrix factorization technique with a neural network to predict criteria evaluation, including the context to predict the overall recommendation. Since expected criteria ratings are inherently uncertain, we also incorporate the Dempster-Shafer assumption function theory into our model to gain criteria rating predictions and develop a discounting and combining scheme to aggregate multiple criteria ratings to obtain an overall rating [2]. The paper by Le et al. served as a particular motivation [14]. The authors used a standard deep learning model (DNN) here. Our goal was to use a graph neural network and to compare both approaches. We show that an approach based on a combination of graph neural networks and belief function theory improves recommender systems' prediction accuracy.
The remainder of this document is organized as follows. Section II briefly recalls some related works, and Section III provides information on the proposed recommendation method. Section IV describes the experimental results and analysis. Finally, Section V concludes the paper with conclusions and future work.

2 Related Work

Since their inception, the collaborative filtering algorithm in recommender systems has been present in e-commerce. The similarity between users or items is used here to predict recommendations [17]. The basis is, therefore, the correlation between users and items. Other algorithms based on data-based model creation are Latent Semantic Analysis [9], Support Vector Machines [21], Bayesian Clustering [4], and Singular Value Decomposition (SVD) [12]. A large number of methods are based on matrix factorization (MF). Dimensional matrices representing interactions between users and items are decomposed into matrices of lower dimensions [16]. Relatively recently, approaches using graphical neural networks (GNN) have appeared. They make it possible to solve tasks that can be structured into a graph [5]. The above GNN-based approaches consider the rating information as the user's opinion on the edges between user nodes and items in a bipartite graph, which can be converted into a general graph.

Some recommendation algorithms consider only the static relationships between the user and the item or even consider time as another parameter. However, there is much more information that must be taken into account when making a recommendation. One of the first works involving multiple parameters can be found in [3]. The work uses collaborative filtering methods, but an aggregation function obtains the overall rating. Other approaches for solving the problem involving additional information, including context, were presented in [15]. If further details such as context and others come into play, a particular aggregation approach needs to be built on, e.g., average, weighted sum, or fuzzy inference system [1, 7, 15, 20] and others [19].

This paper proposes a model including additional information such as context and others. Figure 1 shows the user's interaction with the items with the inclusion of additional information. It can also be represented as a bipartite graph between users and items, with edges labeled with given additional information (e.g., context) and ratings.

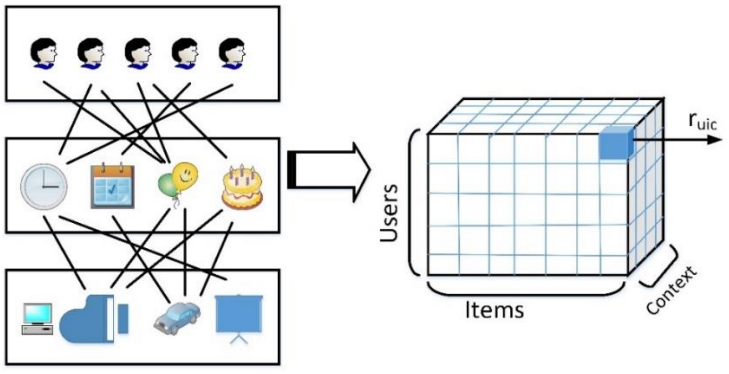


Figure 1 User's interaction with the item (e.g., movie) is surrounded by a specific context (e.g., weather, mood, weekend) that influence the user's opinion on an item. This data forms a 3D matrix between user, item, and context.

3 Methods

We can describe the problem of multicriteria collaborative filtering as follows. We have sets of users, items, and criteria (U, I, C) with set sizes N_U, N_I, and N_C. We will denote the set of multicriteria evaluations as R. These are

data compiled from the assessment of the user *u*, item *i* using criterion *c*. E-commerce systems usually use, for example, a scale of five values {1,2,3,4,5} for presentation {*Horrible*, *Poor*, *Average*, *Good*, *Excellent*}. The set R can be expressed using a 3rd-order tensor. This tensor is usually sparse. Users rate only a limited number of items. The collaborative filtering algorithm has the task of predicting the evaluation of certain unknown items rˆui0 for a particular user so that he can decide, for example, on a purchase, etc.

Our model consists of two main phases:

- Based on the data, the GCN model will be trained to predict the evaluation of individual criteria (e.g., for a film – the performance of the leading actor, etc.), and the parameters for modeling the uncertainty associated with the resulting evaluation will be determined (e.g., the performance of the actor {Terrible, 0.2; Bad, 0.3; Average, 0.5; Good, 0; Excellent, 0}.
- In the second phase, the trained GCN model will be used first to predict the unknown evaluation criteria, which will then serve as sources of evidence represented by assumption function theory (DST) modeling to predict the overall evaluation. A graphical representation of the proposed method is shown in Figure 2.

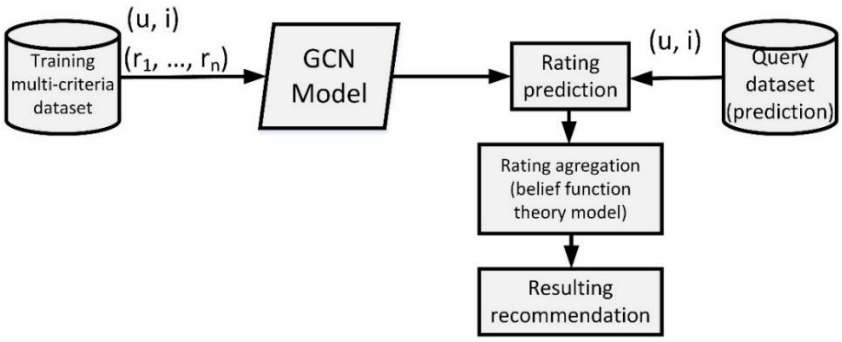


Figure 2 Overview of proposed methods

3.1 Recommendation with GCN

The principle of recommendations based on graph methods is as follows. We have a rating matrix M of $N_u \times N_v$, where N_u is the number of users and N_v is the number of items. The M_{ij} values in this matrix mean either the evaluation performed (user i evaluated item j) from the set of possible evaluation values (discrete) or the fact that the assessment of a particular item j is not performed by user i (value 0 in the M_{ij} matrix). The task of the recommender system is to predict and add missing values to the M_{ij} matrix. So, it indicates how user i would likely rate item j (which he didn't rate). Figure 3 illustrates this situation.

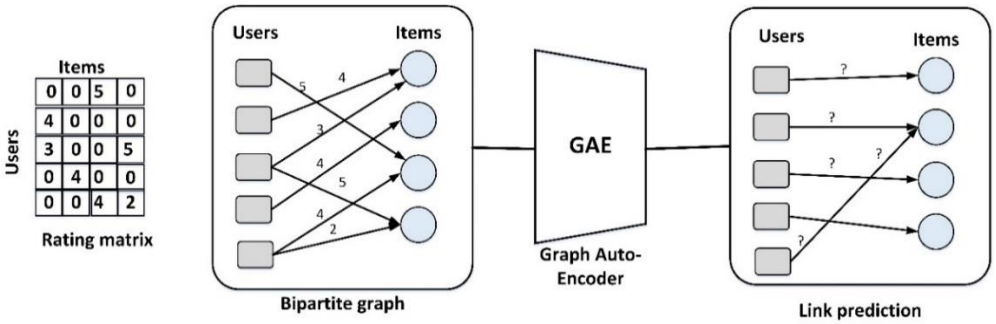


Figure 3 User's interaction and link prediction

The graphical convolutional network (GCN) model starts with an embedding layer in which each user and item is mapped to an *f*-dimensional vector using the SVD method. It is followed by a raft convolutional encoder $[U, V] = f(X, M_1, \ldots, M_R)$ that passes and transforms messages from the user to item nodes and vice versa, followed by a bilinear decoder model that predicts the inputs of the (reconstructed) rating matrix $M' = g(U, V)$, based on user and item input pairs, see Figure 4. Importantly, for our model, the number of neutrons in the output layer of the proposed GCN model must match the number of criteria, which is 5 for our test data set. Other basic optimizers, such as Stochastic Gradient Descent (SGD), can be used to optimize the proposed model.

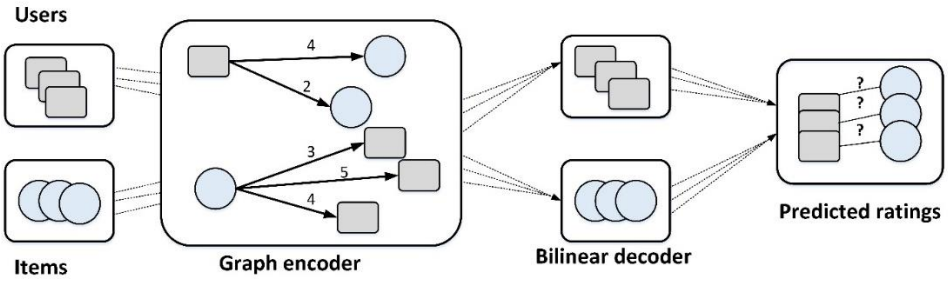


Figure 4 Scheme of the used GCN network

3.2 Evidence Modeling

Information related to decision-making about cyber situations is often uncertain and incomplete. Therefore, finding a feasible way to decide on the appropriate response to the case under this uncertainty is vital. Our model is a particular application of the Dempster-Shafer theory. The Dempster-Shafer theory [18] is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool for combining evidence and changing prior knowledge in the presence of new evidence. The Dempster-Shafer theory can be considered a generalization of the Bayesian theory of subjective probability [18].
Considering a finite set referred to as *the frame of discernment* Θ, a *basic belief assignment (BBA)* is a function $m: 2^{\Theta} \rightarrow [0,1]$ so that

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

where $m(\emptyset) = 0$, see [16]. The subsets of 2^{Θ} which are associated with non-zero values of *m* are known as *focal elements,* and the union of the focal elements is called *the core*. The value of *m(A)* expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Θ belongs to set *A* but not to a specific subset of *A*. This value pertains only to set *A* and makes no additional claims about any subsets of *A*. We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).

Dempster's rule of combination can be used for pooling evidence represented by two values of belief functions over the same frame of discernment coming from independent sources of information. Dempster's rule of combination for combining two belief functions, m_1 and m_2, is defined as follows (the symbol ⊗ is used to denote this operation):

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

where

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

In the proposed recommendation method, the ratings of individual criteria predicted from the trained GCN model are considered evidence for predicting the overall rating. This evidence is represented as mass functions [16] defined by rating grades. Hence, the frame of discernment in our case has five values:

$$\Omega = \{\text{terrible, bad, average, good, excellent}\}.$$

The exact masses m_i of the elements of this frame of discernment are determined based on the probability density function (PDF) estimated using the training data. For each criterion c, we will predict the values of the mass function for individual values of the frame of discernment.
In particular, let $\bar{r}_{uic}$ denote the predicted rating of user u to item *i* in criterion *c*. Then the weight function represents the evidence obtained from the expected rating $\bar{r}_{uic}$, denoted by $m_{\bar{r}_{uic}}: 2^{\Omega} \rightarrow [0,1]$, is defined as [14]:

$$m_{\bar{r}_{uic}}(\{terrible\}) = PDF(\bar{r}_{uic}, \sigma, 1)$$

where $PDF(\bar{r}_{uic}, \sigma, 1$ is the Gaussian probability density function.
Similar to the other rating values of Ω, i.e. {bad, average, good, excellent}.
In the previous step, each predicted criterion rating $\bar{r}_{uic}$ from C is represented by the mass function $m_{\bar{r}_{uic}}$. Then we have the N_C mass functions representing the N_C evidence from the predictions of the $\bar{r}_{uic}$ criteria, for c = 1, . . . , N_C. Consider further that each criterion c is associated with a weight w_c representing its relative importance in contributing to the overall rating. Now we combine these N_C weight functions. $m_{\bar{r}_{uic}}$ considering their relative

importance, create a combined weight function to predict the overall rating. A recommendation can be made within DST using the discount and Dempster rule combination.

4 Experiments and Results

We compared our results using graph neural networks with the results of the publication by Le et al. [14]. The advantage here was that the implementation of the methods described in this work [14] is publicly available on the GitHub repository. It uses the PyTorch library for models using deep neural networks, not graphs. The authors of [14] designated their methods as cGCMC and cGCMC F. To achieve a valid comparison, we used the same procedure as the authors of [14]. We used 60% of the data as the training set, 20% as the validation set, and 20% as the test set for each dataset. We have split the data. Each time the data is shuffled with a different random seed before splitting.
Regarding the dataset, we used the same dataset of 274,572 multicriteria ratings from 84,579 users to 6,854 hotels with a sparsity of 99.9526%. Again, this file is publicly available and is extracted from Tripadvisor.com. The criteria collected are value rating (27.507%), location rating (27.478%), cleanliness rating (27.341%), service rating (7.511%), and overall rating (100%).

Evaluation metrics
For the efficiency measurement, the base loss function and coefficient of determination are adopted as follows [6, 11]:
- Mean Absolute Error (MAE): MAE measures the average size of errors in a set of forecasts.
- Root Mean Square Error (RMSE).

4.1 Performance Comparison

We compared the method proposed in this paper with conventional CF approaches, namely user-based (CF_user) and item-based (CF_item) methods. Furthermore, we compared the results of our approach with the DNN_DST model, which was proposed by LE et al. [14]. Like our model, the DNN_DST model uses the theory of conjecture functions. However, our model uses graph neural networks, while DNN_DST uses deep neural networks. Figure 5 shows that DNN_DST outperforms these two conventional methods regarding recommendation accuracy. Our model is then better than the DNN_DST model. The disadvantage of both models using neural networks is the relatively long calculation time because they need time to train the respective neural network.

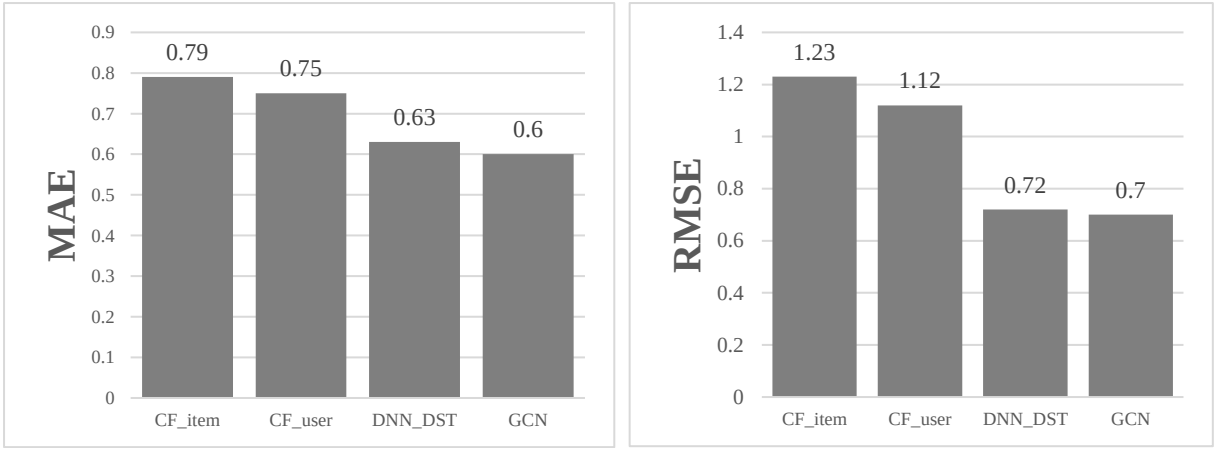


Figure 5 Comparison of our GCN method with other recommendation methods

5 Conclusion

This paper proposes a recommendation calculation using the theory of belief functions and an approach based on graph analysis using graph neural networks. We compared this use of graph neural networks with a similar approach using conjecture function theory with deep neural networks [14].

We proposed a GCN model incorporating the SVD technique as the first layer for prediction. In the next stage, we used an evidential reasoning approach to model these evaluations as evidence using bulk functions. We aggregated to predict the overall rating.
Experiments conducted on a real-world dataset show that the proposed recommendation method outperforms the CF methods in terms of MAE and RMSE scores in most test cases, offering slightly better performance with the technique proposed in the paper [14]. It allows us to state that using graph neural networks gives good results for recommender systems.

References

[1] Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intell. Syst*., 22(3), 48–55.

[2] Beranek, L., Tlusty, P. & Remes, R. (2010). The Usage of Belief Functions for an Online Auction Reputation Model. In *28th International Conference on Mathematical Methods in Economics 2010*, (pp.49-54).

[3] Beranek L & Remes R. (2020). Distribution of Node Characteristics in Evolving Tripartite Network. *Entropy*. 22(3), 263.

[4] Breese, J. S., Heckerman, D. & Kadie, C. (2013). *Empirical analysis of predictive algorithms for collaborative filtering*. [Online]. Available at: https://arxiv.org/ftp/arxiv/papers/1301/1301.7363.pdf.

[5] Cheng, H. T. et al. (2016). Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems* '16. (pp. 7–10).

[6] El Ashmawi, W., Fouad, A. & Slowik, A. (2021). Hybrid crow search and uniform crossover algorithm-based clustering for top-N recommendation system. *Neural Computing and Applications*, 33, 1–20.

[7] He, X., Liao, L., Zhang, H., Nie, L., Hu, X. & Chua, T. (2017). Neural collaborative filtering. *In Proceedings of the 26th international conference on world wide web* (pp. 173–182).

[8] He, M., Zhang, S. & Meng, Q. (2019). Learning to style-aware Bayesian personalized ranking for visual recommendation. *IEEE Access*, 7, 14198–14205.

[9] Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst*., 22(1), 89–115.

[10] Huifeng Guo, H., Tang, R., Ye, Y., Li, Z. & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 1725–1731).

[11] Jannach, D., Karakaya, Z. & Gedikli, F. (2012). Accuracy improvements for multicriteria recommender systems. In *Proc. 13th ACM Conf. Electron. Commerce* (EC), 2012, (pp. 674–689).

[12] Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2008, (pp. 426–434).

[13] Koren, Y., Bell, R. & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Comput*., 42(8), 30–37.

[14] Le, Q. -H., Mau, T. N., Tansuchat, R. & Huynh, V. N. (2022). A Multicriteria Collaborative Filtering Approach Using Deep Learning and Dempster-Shafer Theory for Hotel Recommendations. *IEEE Access*, 10, 37281–37293.

[15] Nassar, N., Jafar, A. & Rahhal, Y. (2020). A novel deep multicriteria collaborative filtering model for recommendation system. *Knowl.-Based Syst*., 187, 104811.

[16] Paterek, A. (2007) Improving regularized singular value decomposition for collaborative filtering. In *Proc. KDD Cup Workshop*, (pp. 5–8).

[17] Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proc. 10th Int. Conf. World Wide Web (WWW)*, 2001, (pp. 285–295).

[18] Shafer, G. (1975). *A mathematical theory of evidence*. Princeton: Princeton University Press.

[19] Stastný, J., Skorpil, V., Balogh, Z. & Klein, R. (2021). Job shop scheduling problem optimization by means of graph-based algorithm. *Applied Sciences*, 11(4), 1921.

[20] Tang T. Y. & McCalla, G. (2009). The pedagogical value of papers: A collaborative-filtering based paper recommender. *J. Digit. Inf.*, 10(2), 1–12.

[21] Wang, R., Cheng, H. K., Jiang Y. & Lou, J. (2019). A novel matrix factorization model for recommendation with LOD-based semantic similarity measure. *Expert Syst. Appl*., 123, 70–81.

[22] Xia, Z., Dong, Y. & Xing, G. (2006). Support vector machines for collaborative filtering. In *Proc. 44th Annu. Southeast Regional Conf.* (*ACM-SE*), (pp. 169–174).

[23] Zhang, S., Yao, L., Sun, A. & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv*., 52(1), 1–38.

Numerical Valuation of Investment Opportunities under Two-Factor Uncertainty

Jiří Hozman[1], Tomáš Tichý[2]

Abstract. Real options approach applies to a wide range of investment opportunities in order to help investors achieve better risk management and more robust financial outcomes. In this paper we focus on a decision-making framework that incorporates two sources of uncertainty in evaluating strategic investments, namely unit output commodity price and unit cost. Incorporating both factors provides a more realistic and accurate approach to evaluating embedded flexibilities, especially in highly uncertain environments.
Using contingent claim analysis, the values of investment opportunities can be identified as solutions to the relevant two-factor Black-Scholes equations, adjusted to match the specific features of real options. As explicit formulae for this kind of PDE problem are only available in certain scenarios (as for conventional financial options), one must rely on numerical techniques in general. Inspired by the methodology from numerical valuation of one-factor real options, we employ and extend the discontinuous Galerkin approach to the two-factor option case exercisable at a fixed time (i.e., European-style option). Finally, the proposed numerical scheme is applied to a simple conceptual expansion decision problem for illustration purposes.

Keywords: real option pricing, project value, option value, two-factor Black-Scholes equation; discontinuous Galerkin method

JEL Classification: C44, G13
AMS Classification: 65M60, 35Q91, 91G60

1 Introduction

Corporate finance has capital budgeting as a one of its crucial disciplines. Its primary objective is to increase the value of the firm via long term investment decisions. Assessing the profitability of investment opportunities is therefore of considerable importance.

In the late 1970s, the foundation of contemporary investment theory was established, see the groundbreaking paper by Myers [13]. This theory links the evaluation of corporate investment flexibilities to the pricing of financial options on tangible assets, forming what is now commonly known as the real options approach. This methodology treats the value of an investment opportunity as a similar option premium on a financial asset. Since its inception, many solution techniques have emerged, including simulation approaches, dynamic programming, and contingent claims analysis, for a brief overview see [12]. Among them, the most developing technique belongs to the contingent claims framework, which allows us to formulate a real option pricing problem through a partial differential equation (PDE) and, secondly, to use modern numerical approaches.

In this short contribution we extend our recent results focused on the numerical valuation of real options under one-factor uncertainty (in commodity price), where the discontinuous Galerkin (DG) method is applied to solve the similar governing equations, see conference papers [9] and [10]. We proceed as follows – in Section 2 the relevant PDE models of two-factor uncertainty, arising from [2], are formulated, while in Section 3 a numerical approach is presented. Finally, in Section 4 a simple conceptual experiment related to a decision problem of doubling production is provided.

2 Real Options with Two-Factor Uncertainty

The real options approach applied within a finite time horizon employes a backward induction procedure to evaluate the embedded flexibility of an investment opportunity described by relevant PDEs. In contrast to one-factor stochastic framework from [9], we consider here a risk-neutral firm (e.g., mining company) that has the opportunity to invest in a (mining) plant which is characterized by stochastic cost per unit C and stochastic revenue

[1] Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz
[2] Department of Finance, VSB-TU Ostrava, Sokolská třída 33, 702 00, Ostrava, Czech Republic, tomas.tichy@vsb.cz

determined by the output commodity price per unit P. Both price P and cost C are modeled as processes following geometric Brownian motion (as proposed in [17]):

$$\mathrm{d}P(t) = \alpha_P P(t)\mathrm{d}t + \sigma_P P(t)\mathrm{d}W_P(t), \quad P(0) > 0, \tag{1}$$

$$\mathrm{d}C(t) = \alpha_C C(t)\mathrm{d}t + \sigma_C C(t)\mathrm{d}W_C(t), \quad C(0) > 0, \tag{2}$$

where t is the time component, α_P, α_C are growth rates, $\mathrm{d}W_P$, $\mathrm{d}W_C$ increments of the Wiener processes, and $\sigma_P > 0$ is the volatility of the output commodity price and $\sigma_C > 0$ is the volatility of the cost process, respectively. Usually, $\alpha_P = r - \delta$, where r is the risk-free interest rate and $\delta > 0$ is the mean convenience yield on holding one unit of the output. More specifically, we allow the processes (1) and (2) to be correlated by the factor $\rho \in (-1, 1)$. This simultaneous presence of two stochastic components represents an essential extension of the methodological concepts of real options valuation from [9] which involves adjusting the governing equations, see paragraphs below.

The problem is to find the true value of the investment opportunity, $F(P, C, t)$, arising from values of the firms $V_0(P, C, t)$ and $V_1(P, C, t)$, i.e., having no investment opportunity and the embedded one allowing a single decision-making at a fixed $T > 0$, respectively. Based on the contingent claims analysis [1], value functions V_i, between the change of operating T and the project life-time T_i^*, must satisfy the following PDEs (cf. [4]):

$$\frac{\partial V_i}{\partial t} + \mathcal{L}_{\mathrm{BS}}(V_i) = -\varphi_i \quad \text{in } (0,\infty)^2 \times [T, T_i^*),\ i = 0, 1, \tag{3}$$

where the differential operator $\mathcal{L}_{\mathrm{BS}}$ is of the Black-Scholes (BS) type, defined as

$$\mathcal{L}_{\mathrm{BS}}(V_i) = \frac{1}{2}\sigma_P^2 P^2 \frac{\partial^2 V_i}{\partial P^2} + \rho\sigma_P\sigma_C PC\frac{\partial^2 V_i}{\partial P \partial C} + \frac{1}{2}\sigma_C^2 C^2 \frac{\partial^2 V_i}{\partial C^2} + \alpha_P P\frac{\partial V_i}{\partial P} + \alpha_C C\frac{\partial V_i}{\partial C} - rV_i \tag{4}$$

and $\varphi_i(P, C, t)$, $i = 0, 1$, represent (after-tax) cash flow rate associated with the investment opportunity. Since the value of the firm is worthless at $t = T_i^*$, we subject (3) to the terminal state $V_i(P, C, T_i^*) = 0$, $i = 0, 1$.

Further, at $t = T$, the company values and the value of the investment opportunity are interconnected by the following relationship

$$F(P, C, T) = \Pi(P, C) \equiv \max\Big(V_1(P, C, T) - V_0(P, C, T) - \mathcal{K}, 0\Big), \quad [P, C] \in (0,\infty)^2, \tag{5}$$

where $\mathcal{K}$ is implementation costs (if positive) or disinvestment costs (if negative).

Finally, we define the value function $F(P, C, t)$ as the difference $V_1(P, C, t) - V_0(P, C, t)$ for all $t \in [0, T)$, that represents the value added to the value of the firm. Within the timeline $[0, T)$, taking into account an equivalence of cash flow rates, the value function F satisfies the governing equation with the same differential operator as in (3), but with a zero right-hand side, i.e.,

$$\frac{\partial F}{\partial t} + \mathcal{L}_{\mathrm{BS}}(F) = 0, \quad \text{in } (0,\infty)^2 \times [0, T). \tag{6}$$

The resulting governing equations (3) and (6) belong to the class of convection-diffusion problems, which displays both parabolic and hyperbolic qualities depending on the BS model parameters. Therefore, the DG technique is implemented to address this issue, which has been effective in pricing conventional financial options, see, e.g. [7] and [8].

3 Numerical Approach

Real options valuation relies heavily on modern numerical methods, as analytical formulae for BS type PDEs are readily available only in the simplest scenarios or in very limited situations. The proposed valuation methodology, based on the DG method, constructs a numerical solution as a composition of piecewise polynomial, generally discontinuous, functions on finite element mesh without any requirements on the continuity of the solution across the elements, see [15].

This numerical approach leads to the discretization of the spatial domain that has to be bounded. Therefore, we restrict the governing equations and the relevant terminal conditions to the bounded (P, C)-domain. For this purpose let $P_{\max}$ and $C_{\max}$ denote the maximal sufficient value of the unit commodity price and maximal possible value of unit cost, respectively. Without loss of generality $C_{\max} = P_{\max}$, i.e., we consider the square domain $\Omega = (0, P_{\max})^2$. The firm as well as investment opportunity valuation problems are newly represented by the

initial–boundary value ones, which must be equipped with additional boundary conditions. We follow the common approach, where the boundary conditions are chosen in accordance with the algebraic sign of the so called Fichera function defined on $\partial\Omega$, for a detailed explanation see [14]. As a result, to guarantee the well-posedness, no boundary conditions have to be imposed on $P = 0$ and $C = 0$. On the other hand, we prescribe the artificial boundary conditions on the far-field boundary (i.e., $P = P_{\max}$ and $C = C_{\max}$) that are consistent with terminal states and respect the asymptotic behaviour given by net present value rule and conventional financial options, see [11]. Accordingly, we prescribe

$$V_i(P_{\max}, C, t) = \int_t^{T_i^*} \varphi_i(P_{\max}, C, t) e^{-r(\xi - t)} \mathrm{d}\xi, \quad C \in [0, C_{\max}], \quad t \in [T, T_i^*), \ i = 0, 1, \tag{7}$$

$$V_i(P, C_{\max}, t) = \int_t^{T_i^*} \varphi_i(P, C_{\max}, t) e^{-r(\xi - t)} \mathrm{d}\xi, \quad P \in [0, P_{\max}], \quad t \in [T, T_i^*), \ i = 0, 1 \tag{8}$$

and

$$F(P_{\max}, C, t) = e^{-r(T-t)} \Pi(P_{\max}, C), \ C \in [0, C_{\max}], \ t \in [0, T), \tag{9}$$

$$F(P, C_{\max}, t) = e^{-r(T-t)} \Pi(P, C_{\max}), \ P \in [0, P_{\max}], \ t \in [0, T). \tag{10}$$

The whole discretization procedure, with respect to the space-time domain of governing equations, consists of two consecutive phases — spatial semi-discretization and temporal discretization, see [7]. Respecting this, it is necessary to proceed by backward induction to determine the value of the investment opportunity, i.e., from a pair of firm value functions $V_0(P, C, t)$ and $V_1(P, C, t)$, over a construction of the interconnecting function $\Pi(P, C)$ to the desired real option value, $F(P, C, 0)$.

Let $S_h^p(\Omega)$ be a finite dimensional space consisting of piecewise polynomial, generally discontinuous, functions of the p-th order defined over the triangulation of the domain Ω with the assigned mesh size h. We define by $u_h^{(i)}(t) \in S_h^p(\Omega)$, $t \in [T, T_i^*]$, $i = 0, 1$, the DG semi-discrete functions related to the values of firms $V_i(\cdot, \cdot, t)$, $i = 0, 1$. Similarly, $w_h(t) \in S_h^p(\Omega)$, $t \in [0, T]$ represents the semi-discrete version of $F(\cdot, \cdot, t)$. The resulting numerical procedure can be summarized into the following steps:

(S1) Set $u_h^{(i)}(T_i^*) = 0, i = 0, 1$.

(S2) The functions $u_h^{(i)}(t)$, $t \in [T, T_i^*)$, $i = 0, 1$, are defined as solutions of systems of ordinary differential equations (ODEs), using a variational formulation similar to [6], i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(u_h^{(i)}(t), v_h\right) + \mathcal{D}_h^{\mathrm{DG}}\left(u_h^{(i)}(t), v_h\right) = \ell_h^{(i)}(v_h)(t) - \left(\varphi_i(t), v_h\right) \quad \forall\, v_h \in S_h^p(\Omega), \forall\, t \in [T, T_i^*), \quad i = 0, 1, \tag{11}$$

where $(\cdot, \cdot)$ denotes the inner product in $L^2(\Omega)$, the bilinear form $\mathcal{D}_h^{\mathrm{DG}}(\cdot, \cdot)$ stands for the DG semi-discrete variant of (4) and the linear form $\ell_h^{(i)}(\cdot)(t)$, $i = 0, 1$, enforces the boundary conditions (7) and (8), see [7] for details.

(S3) Set the terminal condition for the single real option by S_h^p-projection, i.e.,

$$\left(w_h(T), v_h\right) = \left(\max\left(u_h^{(1)}(T) - u_h^{(0)}(T) - \mathcal{K}, 0\right), v_h\right) \quad \forall\, v_h \in S_h^p(\Omega). \tag{12}$$

(S4) Similarly to (S2), the function $w_h(t)$, $t \in [0, T)$, is defined as follows

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(w_h(t), v_h\right) + \mathcal{D}_h^{\mathrm{DG}}\left(w_h(t), v_h\right) = \ell_h(v_h)(t) \quad \forall\, v_h \in S_h^p(\Omega), \forall\, t \in [0, T), \tag{13}$$

where the linear form $\ell_h(\cdot)(t)$, $i = 0, 1$ balances boundary conditions (9) and (10).

Finally, note that problems (11) and (13) have to be discretized with respect to the time coordinate to obtain a fully discrete solutions on time levels $t = T$ and $t = 0$, respectively. As a finite difference scheme we use an implicit Euler method, which has no restrictive condition on the length of the time step. The result is a sequence of systems of linear algebraic equations with sparse non-symmetric matrices that are treated by a restarted GMRES solver within the numerical procedure, see [8] for details.

4 Conceptual Experiment of Expansion Option

This section provides a brief demonstration of the application of the numerical approach in a conceptual study related to the iron ore mining sector. We consider a one-stage investment project to double the production of an

iron ore, arising from [11], where the case of the commodity price uncertainty was investigated. Although the following example is illustrative in nature, it is related to relevant data of practical significance and allows for proper transferability to real-world case studies. The output commodity price P and mining cost C are expressed in USD per dry metric tonne (dmt) of iron ore. The considered investment opportunity F is represented by a (real) option to expand the production, exercisable at fixed time $T = 2$ (in years) and requiring the investment cost $\mathcal{K} = 10$ (in thousands of million USD). From the point view of conventional financial options such an investment opportunity can be interpreted as a call option under the European exercise right with strike $\mathcal{K}$ and maturity date T.

Let $Q = 10$ be the total reserve of the iron ore mine (in thousands of million dmt) and $q_0(t)$, $q_1(t)$ be the iron ore production rates (in thousands of million dmt per year) associated with the value of the mine itself (V_0) and the mining company with the embedded investment opportunity (V_1), respectively. Depending on how the mine is operated, project lifetimes are defined as minimum admissible values T_0^* and T_1^* (in years) that satisfy the relationship

$$Q = \int_0^{T_0^*} q_0(\xi)\mathrm{d}\xi = \int_0^{T_1^*} q_1(\xi)\mathrm{d}\xi, \tag{14}$$

$$q_0(t) = 0.1\, e^{0.007t}, \quad t \in [0, T_0^*), \qquad\qquad q_1(t) = \begin{cases} q_0(t), & \text{if } t \in [0, T), \\ 2 \cdot q_0(t), & \text{if } t \in [T, T_1^*). \end{cases} \tag{15}$$

Further, we define the after-tax cash flow rates of relevant projects as modification from [11] in the following form

$$\varphi_i(P, C, t) = q_i(t)\Big((1 - D)P - C\Big)(1 - B), \quad [P, C] \in \overline{\Omega},\ t \in [0, T_i^*], \qquad i = 0, 1, \tag{16}$$

where D is the rate of state royalties and B the income tax rate, respectively. The numerical experiments are performed on the following (reference) market data: $D = 0.05$, $B = 0.30$, $r = 0.06$, $\alpha_P = 0.04$, $\alpha_C = 0.005$, $\sigma_P = 0.35$, $\sigma_C = 0.35$, $\rho = 0.40$, which are related to [11].

Referring to [9] we price the corresponding real option using a piecewise quadratic approximation (i.e., $p = 2$) on the fixed uniformly partitioned square grid of domain Ω with the unit mesh size (i.e., $h = 1$) and $P_{\max} = C_{\max} = 100$. In parallel with this, the time step is set as $\tau = 0.02$. Using (14) and (15), easy calculation leads to $T_0^* \doteq 75.8$ and $T_1^* \doteq 43.6$, respectively. The valuation scheme (S1)–(S4) is implemented in the solver Freefem++, see [5].

First of all, we want to graphically capture a final numerical solution in the whole computational domain Ω, see Figure 1. The nature of the numerical solution is significantly influenced by the shape of the cash flow rate (16). Since $\varphi_i(P, C, t)$ is constant across lines $0.95P - C = const.$, this property is partially inherited into the resulting real option value. However, the solution also takes into account the direction of the vector field given by the convection fluxes

$$\Big((\alpha_P - \sigma_P^2 - \rho\sigma_P\sigma_C/2)P,\ (\alpha_C - \sigma_C^2 - \rho\sigma_P\sigma_C/2)C\Big), \quad [P, C] \in \overline{\Omega}. \tag{17}$$

As (17) is proportional to P, C and time runs backwards, the value from the far-field boundary is propagated to the zone of interest as $t \to 0+$, or vice versa depending on the sign of convection fluxes. Simultaneously, the contribution of the diffusion term is significant due to a relatively high volatile scenario ($\sigma_P = \sigma_C = 0.35$) and thus the solution is smoothed as it is developed from the stage $\Pi(P, C)$. The whole (aforementioned) behaviour is better observable on plots with isolines of solutions, see Figure 1 (left). Moreover, one can easily recognize that a shape of a real option value function along line $C = const.$ is similar to the conventional financial European call option with the relevant BS parameters r and σ_P.

Secondly, our aim is to provide the deeper insight to the sensitivity measures of the studied real option valuation problem. Therefore, we introduce the Delta measures $\Delta_h^P(t) \approx \frac{\partial F}{\partial P}(\cdot, \cdot, t)$ and $\Delta_h^C(t) \approx \frac{\partial F}{\partial C}(\cdot, \cdot, t)$ whose present values (at $t = 0$) are depicted in Figure 2 and Figure 3, respectively. At first glance, the most sensitive real option value with respect to the output commodity price is related to the region of low values of mining cost, see Figure 2. This phenomenon corresponds to the intuitive expectation that value of investment opportunity is strongly influenced by the output price P when mining cost follows $C \to 0+$. On the other hand, Figure 3 (left) reveals that values of investment opportunities dramatically change with respect to the mining cost in the band of the output prices $18 \le P \le 38$ which coincide with experiment setting from [11]. A summary of these graphs also shows that the value of investment opportunity is an increasing function of the output commodity price P and a decreasing one of the mining cost C, concurrently. These observations are fully in line with the expectations of decision makers. Regardless, it should be noted that extreme overshoots in Delta measures along lines $C = 0$ and $P = P_{\max}$ can be eliminated by using proper mesh adaptability [8], which is not used here to ensure consistency with [9]. At this point, it should be also emphasized that the aim of this simple conceptual experiment is to substantiate the design of the numerical scheme as the whole, and we are aware that the presented results are preliminary and more thorough numerical analysis is needed, but it would be beyond the scope of this contribution.

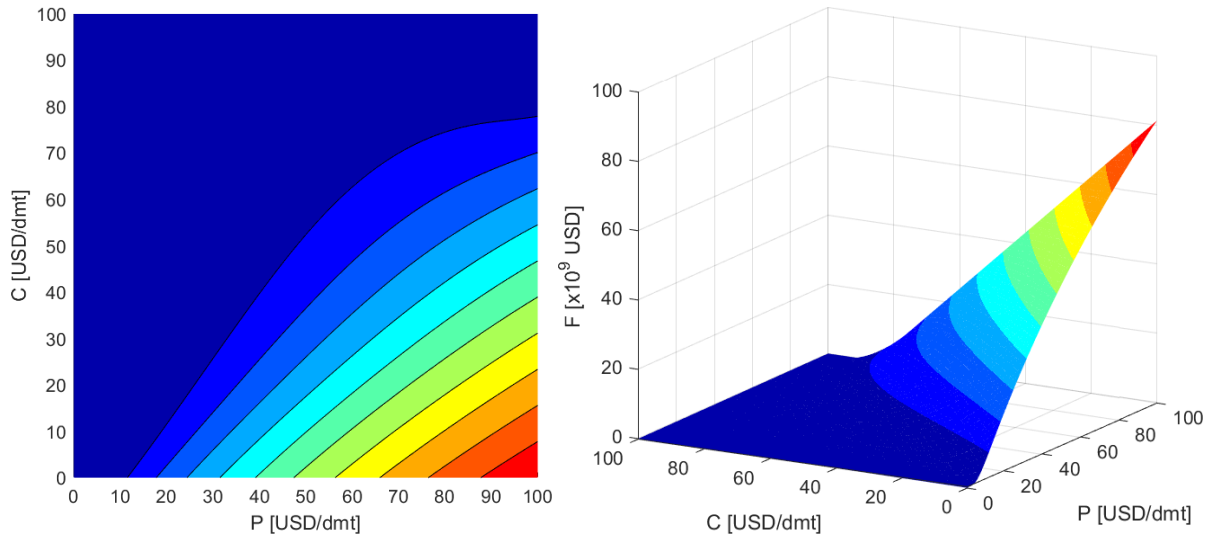


Figure 1 Piecewise quadratic approximation of the value function of the investment opportunity at present time $t = 0$: isoline plot (left) and 3D plot (right)

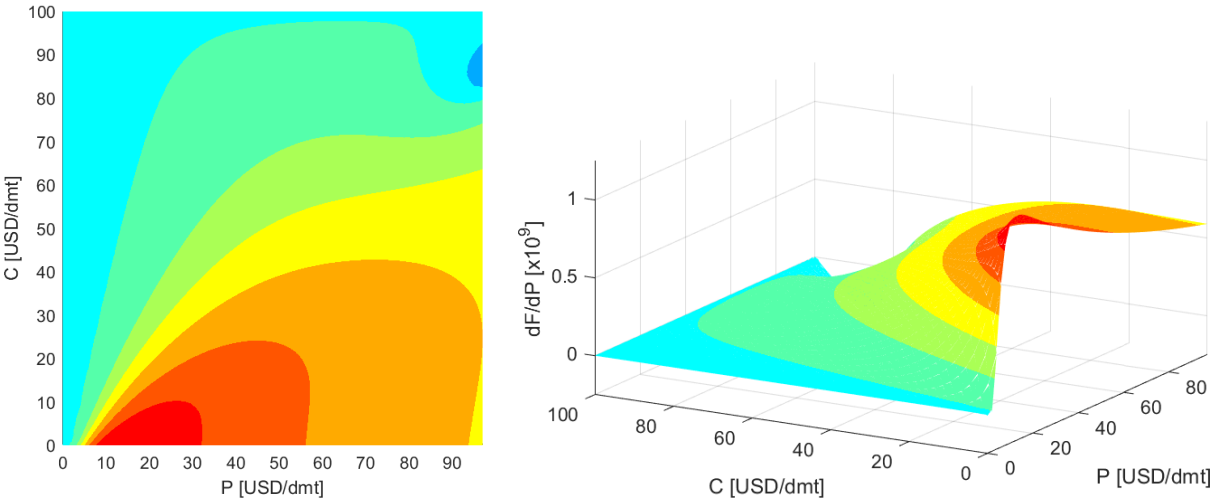


Figure 2 Piecewise linear approximation of the Delta sensitivity measure with respect to P-variable at present time $t = 0$: isoline plot (left) and 3D plot (right)

5 Conclusion

Valuing real options is a critical and complex aspect of decision-making process, and the use of numerical techniques, especially the PDE approach, is crucial in the modern investment framework. The presented methodological concept forms a general two-factor problem to evaluate the investment opportunities, and the proposed numerical scheme is designed to cope with robust scenarios given by a wide range of BS model parameters. The elaborate experiment (performed in the conceptual study) offers a brief understanding of the performance of the numerical scheme and its connection to the decision-making process through financially relevant results. One direction of future research is a multi-stage sequential decision process that at a given stage contains different investment opportunities (as in [16]) that are folded into one compound option, capable of adapting to evolving investment strategies over a longer time frame. Moreover, this is strengthened if a stochastic scenario for the BS model parameters is considered simultaneously.

Acknowledgements

Both authors were supported through the Czech Science Foundation (GAČR) under project 22-17028S. The support is greatly acknowledged. Furthermore, the second author also acknowledges the support provided within SGS 2023/19, a research project of VSB-TU Ostrava.

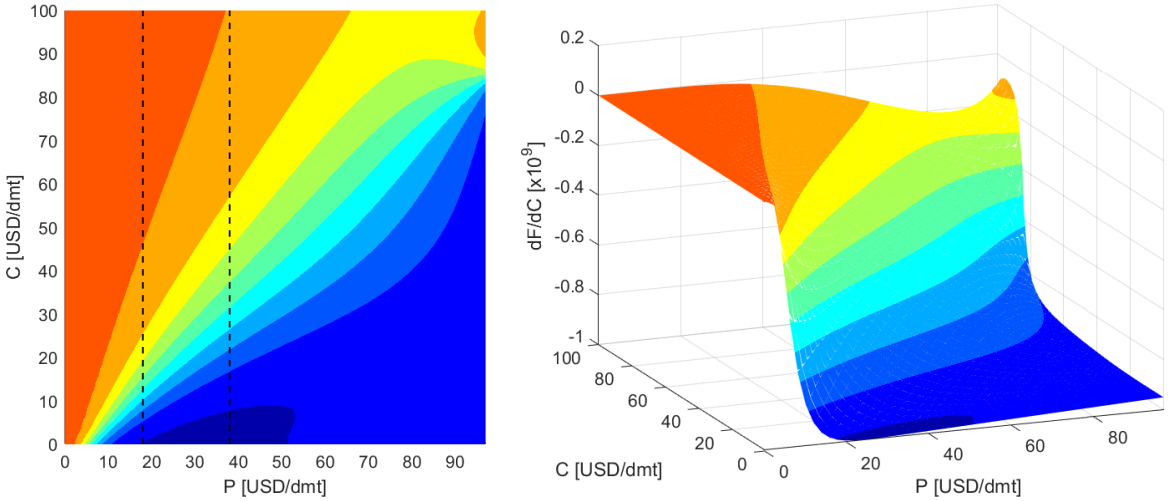


Figure 3 Piecewise linear approximation of the Delta sensitivity measure with respect to C-variable at present time $t = 0$: isoline plot (left) and 3D plot (right)

References

[1] Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–659.

[2] Compernolle, T., Huisman, K.J.M., Kort, P.M., Lavrutich, M., Nunes, C. & Thijssen, J.J.J. (2021). Investment Decisions with Two-Factor Uncertainty. *Journal of Risk and Financial Management*, 14, 534.

[3] Dixit, A. & Pindyck, R. (1994). *Investment Under Uncertainty*. Princeton: Princeton University Press.

[4] Haque, M., Topal, E. & Lilford, E. (2014). A numerical study for a mining project using real options valuation under commodity price uncertainty. *Resources Policy*, 39, 115–123.

[5] Hecht, F. (2012). New development in FreeFem++. *Journal of Numerical Mathematics*, 20, 251–265.

[6] Hozman, J. & Tichý, T. (2015). A discontinuous Galerkin method for pricing of two-asset options. In D. Martinčík (Eds.), *Proceedings of the 33rd International Conference Mathematical Methods in Economics* (pp. 273–278). Plzeň: University of West Bohemia.

[7] 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.

[8] Hozman, J. & Tichý, T. (2020). The discontinuous Galerkin method for discretely observed Asian options. *Mathematical Methods in the Applied Sciences*, 43, 7726–7746.

[9] Hozman, J. & Tichý, T. (2021). Numerical Valuation of the Investment Project with Expansion Options Based on the PDE Approach. In R. Hlavatý (Eds.), *Proceedings of the 39th International Conference Mathematical Methods in Economics* (pp. 185–190). Prague: Czech University of Life Sciences Prague.

[10] Hozman, J. & Tichý, T. (2022). Numerical Valuation of the Investment Project Flexibility Based on the PDE Approach: An Option to Contract. In H. Vojáčková (Eds.), *Proceedings of the 40th International Conference Mathematical Methods in Economics* (pp. 122–128). Jihlava: College of Polytechnics Jihlava.

[11] Li, N. & Wang, S. (2019). Pricing options on investment project expansions under commodity price uncertainty. *Journal of Industrial & Management Optimization*, 15, 261–273.

[12] Mun, J. (2002). *Real Options Analysis: Tools and Techniques for Valuing Strategic Investments and Decisions*. John Wiley & Sons, Inc., Hoboken.

[13] Myers, S.C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5, 147–175.

[14] Oleinik, O.A. & Radkevič, E.V. (1973). *Second Order Equations with Nonnegative Characteristic Form*. Boston: Springer-Verlag.

[15] Rivière, B. (2008). *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. SIAM, Philadelphia.

[16] Trigeorgis, L. (2005). Making use of real options simple: An overview and applications in flexible/modular decision making. *The Engineering Economist*, 50, 25–53.

[17] Tschulkow, M., Compernolle, T. & Van Passel, S. (2021). Optimal timing of multiple investment decisions in a wood value chain: A real options approach. *Journal of Environmental Management*, 290, 112590.

Cost Optimization Model for Synchronous Storage and Dispatch in Production Warehouse

Andrea Hrníčková[1], Karel Ječmen[2], Dušan Teichmann[3], Denisa Mocková[4],

Abstract. There are various types of warehouses in the logistics chain, one of them are production warehouses. These are warehouses, which storage capacity is used to complete product processes before shipping to the customer or to homogenize the qualitative parameters of the material before processing. This process is called refinement. Typical examples of products that go through the refining process can be found in the food and chemical industry, especially fermentation processes in brewing and distilling, or stabilizing chemical processes in cosmetology. These products usually require transportation to the production warehouse immediately after production, where take place storage with a refining process while observing specific, predefined conditions (for example, a certain temperature or humidity in the warehouse). In these conditions, the products must spend the minimum required time, and in most cases the stay in the warehouse must not exceed the maximum defined time, otherwise the final product could deteriorate. These mentioned factors affect not only the quality of the product, but also the financial complexity of the entire process. This paper deals with the application of the linear programming method for the purpose of optimizing the total costs of production, synchronous warehousing process with the refining process and dispatching finished products to customer using a heterogeneous vehicle fleet.

Keywords: costs optimization, linear mathematical programming, production warehouse

JEL Classification: C61
AMS Classification: 90C15, 90C08

1 Introduction

The paper thematically follows the presented paper at the Strategic Management and its Support by Information Systems 2019 "Model of storage and shipping synchronization in production warehouses". The paper dealt with the creation of a mathematical model of a production warehouse with a focus on the temporal coordination of the processes of warehousing and dispatch of products while defining the minimum and maximum duration of a product's stay in a production warehouse using a homogeneous fleet of vehicles. The task was solved for the observed planning period consisting of 8 intervals and a production volume of size Q to be produced was given. For each interval, the loading capacity CN_i, the production warehouse capacity CS_i were known, and the large production volume is limited by the available production capacities a_i, which the warehouse can accept for refining. For refined products, the minimum time $MINIM$ and the maximum time $MAXIM$, which they should be stored in the production warehouse, were known. Transportation to the main customer was ensured using a homogeneous vehicle park with a vehicle capacity of K. The task was to determine the volumes of additional production in each interval so that the production was dispatched to the customer by the minimum number of vehicles.

In [4] the optimization criterion is the total number of vehicles intended for the dispatch of additionally produced production in each interval, and the goal of the optimization process is its minimization. In the presented paper,

[1] Czech Technical University in Prague, Department of Logistics and Management of Transport, Horská 3 128 03 Praha 2, hrnicand@fd.cvut.cz.
[2] Czech Technical University in Prague, Department of Logistics and Management of Transport, Horská 3 128 03 Praha 2, jecmekar@fd.cvut.cz.
[3] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu 15/2172, 708 33, Czech Republic, dusan.teichmann@vsb.cz.
[4] Czech Technical University in Prague, Department of Logistics and Management of Transport, Horská 3 128 03 Praha 2, mockova@fd.cvut.cz.

the optimization criterion will be the total costs of the process of production, synchronous warehousing and refining, and dispatching. The distribution of the finished product to the customer will be carried out using a heterogeneous vehicle fleet. Unit costs for production, storage, refining and handling will be newly included in the model. The model also includes vehicle loading costs, which are proportional to its size. The other quantities remain the same as in the paper [4].

2 Motivation to Solve the Problem and Analysis of the Current State of Knowledge

Optimizing the synchronous warehousing process with subsequent dispatch in production warehouses is a process influenced not only by production, its procedures, and outputs, but also by customer requirements, which may also be of a seasonal nature.

Thesis [1] describes in detail the process of beer distribution from the producer through distribution warehouses to wholesale customers. The problem is solved by a mathematical linear model, the optimization criterion are the total costs of distribution.

In [3], a decision support system for multi-level production planning, which was implemented in the Swiss brewery Feldschlösschen, is presented. The support system is based on mathematical optimization. The brewing system in the brewery contains 13 production and 8 storage resources, produces 220 finished products and 100 semi-finished products, and the planning horizon can exceed up to 52 weeks. Paper [6] also deals with solving of the problem of planning beer production by linear mathematical programming. The proposed model is solved by the CPLEX solver. Unlike the previous paper, it deals with the production of a smaller size.

In [2], a new MILP-based solution framework is presented for the problem of optimal planning and scheduling of production in breweries with a large amount of production. The production process is divided into a batch phase, which includes the refining process, and a continuous phase, where the finished product is bottled. This is a complex problem, for which a long planning horizon must be considered due to the ongoing maturation of the beer, and attention must be paid to the correct synchronization of the individual stages of the production process. In [5], a Mixed-Integer Linear Programming model (MILP) was presented for planning and scheduling the production of syrups and jams under the conditions of a small manufacturer. The specific of this manufacturer is that several production lines for the production of syrups and jams share the same fruit hopper with a limited capacity. The model finds a strategy for synchronizing the production process of products with usable capacity leading to profit maximization.

3 Problem Formulation and its Mathematical Model

In this chapter, the solved problem is presented and a mathematical model, which is used to perform the experiments described in the following chapter is presented.

3.1 Formulation of Optimization Problem

A planning period is given, in which a certain volume of homogeneous production of size Q is to be manufactured and dispatched. The planning period is divided into n time intervals. For each interval $i = 1, \ldots, n$, the volume of production that can be delivered to the warehouse is known (it is limited by the available production capacities a_i), the loading capacity CN_i and the production warehouse capacity CS_i are known (these values are always detected at the beginning of the time interval). Also known are the costs of producing one unit d, the costs of the synchronous storage process with the refining process of one unit of production f, and the costs of handling one unit e. Furthermore, a minimum time $MINIM$ and a maximum time $MAXIM$, which the products must spend in production warehouse, are known. Assume that the $MINIM$ and $MAXIM$ values are constant during the whole planning period. The newly produced production is taken by 1 customer, while the transport to this customer can be realized in vehicle types, while for each vehicle type $l = 1, .., v$ the capacity K_l and the cost of accepting the vehicle for loading N_l are given. The task is to determine the volumes of additional production in individual intervals so that the total costs of production, storage with the refining process and dispatching are minimal and at the same time the customer's additional requirements are satisfied. Considering the required minimum storage time, it makes sense to consider capacities a_i only for $i = 1, \ldots, n - MINIM$. If the time of stay in the warehouse is not limited from below, $MINIM = 0$ and the newly arrived order can be realized in every time interval of the planning period.

In the task, let's further assume that at the beginning and at the end of the planning period, the warehouse must be empty in terms of the additional production volume. A zero inventory level in the warehouse at the beginning of the planning period means that there is no free production available to fulfil the new customer's request. A zero inventory level at the end of the planning period means that we do not have any quantity of the manufactured production in stock to cover additional extraordinary customer requirements arising after the end of the planning period.

In order to model the decision, we introduce the following groups of variables into the problem:

$\boldsymbol{x_i}$	the amount of production stored in the time interval $i = 1, \dots, n - MINIM$
$\boldsymbol{y_{jl}}$	number of vehicles of type $l = 1, \dots, v$ loaded in the time interval $j = 1 + MINIM, \dots, n$
$\boldsymbol{w_i}$	the volume of production in the warehouse in the time interval $i = 0, \dots, n$ (the amount shipped in the time unit following the time point $i = 1, \dots, n$ is not included)
$\boldsymbol{z_{ijl}}$	the amount of production delivered to the warehouse in the time interval $i = 1, \dots, n - MINIM$ and dispatched by vehicle type $l = 1, \dots, v$ in the time interval $j = 1 + MINIM, \dots, n$

Table 1 List of the variables

The mathematical model has the form:

$$min\, f(x,y,z,w) = \sum_{i=1}^{n-MINIM} x_i \cdot d + \sum_{i=0}^{n} w_i \cdot f + \\ + \sum_{i=1}^{n-MINIM} \sum_{j=1+MINIM}^{n} \sum_{l=1}^{v} z_{ijl} \cdot e + \sum_{j=1+MINIM}^{n} \sum_{l=1}^{v} y_{jl} \cdot N_l \tag{1}$$

subject to:

$$\sum_{i=1}^{n-MINIM} x_i = Q \tag{2}$$

$$x_i \le a_i \qquad i = 1, \dots, n - MINIM \tag{3}$$

$$x_i = \sum_{j=i+MINIM}^{i+MAXIM} \sum_{l=1}^{v} z_{ijl} \qquad i = 1, \dots, n - MAXIM \tag{4}$$

$$x_i = \sum_{j=i+MINIM}^{n} \sum_{l=1}^{v} z_{ijl} \qquad i = n - MAXIM + 1 \dots n - MINIM \tag{5}$$

$$\sum_{i=1}^{j-MINIM} z_{ijl} \le K_l y_{jl} \qquad \begin{matrix} j = 1 + MINIM, \dots, 1 + MAXIM; \\ l = 1, \dots, v \end{matrix} \tag{6}$$

$$\sum_{i=j-MAXIM}^{n-MINIM} z_{ijl} \le K_l y_j \qquad j = 2 + MAXIM, \dots, n;\ l = 1, \dots, v \tag{7}$$

$$w_0 = 0 \tag{8}$$

$$w_n = 0 \tag{9}$$

$$w_j = w_{j-1} + x_j \qquad j = 1..MINIM \tag{10}$$

$$w_j = w_{j-1} + x_j - \sum_{i=1}^{j-MINIM} \sum_{l=1}^{v} z_{ijl} \qquad j = MINIM + 1, \dots, MAXIM \tag{11}$$

$$w_j = w_{j-1} + x_j - \sum_{i=j-MAXIM}^{j-MINIM} \sum_{l=1}^{v} z_{ijl} \qquad j = MAXIM + 1, \dots, n - MINIM \tag{12}$$

$$w_j = w_{j-1} - \sum_{i=j-MAXIM}^{j-MINIM} \sum_{l=1}^{v} z_{ijl} \qquad j = n - MINIM + 1, \dots, n \tag{13}$$

$$w_i \le CS_i \qquad i = 1, \dots, n \tag{14}$$

$$\sum_{l=1}^{v} y_{jl} \leq CN_j \qquad j = 1 + MINIM, \dots, n \qquad (15)$$

$$x_i \geq 0 \qquad i = 1, \dots, n - MINIM \qquad (16)$$

$$y_{jl} \in Z_0^+ \qquad j = 1 + MINIM, \dots, n;\ l = 1, \dots, v \qquad (17)$$

$$w_i \geq 0 \qquad i = 0, \dots, n \qquad (18)$$

$$z_{ijl} \geq 0 \qquad i = 1, \dots, n - MINIM; j = 1 + MINIM, \dots, n;\ l = 1, \dots, v \qquad (19)$$

Function (1) represents the optimization criterion – the total cost of the process. In the total costs, the first part corresponds to production costs, the second part to storage and refining costs, and the third part is made up of handling costs and the costs of receiving the vehicle for loading, i.e., the costs of dispatching the products to the customer after the refining phase. Constraint (2) ensures that the production volume requested by the newly arrived customer will be realized. A group of constraints (3) ensures compliance with the free capacity of production in each time interval. Groups of constraints (4) and (5) ensure the implementation of the refinement phase. This means that production will be spread over intervals defined by the minimum and maximum limits of the refinement process. Groups of constraints (6) and (7) ensure that the conversion of the dispatched quantity in each time interval to the number of vehicles of individual types, which will be loaded and dispatched to the customer in these intervals. Constraints (8) and (9) ensure that at the beginning and at the end of each interval, there will be an available amount of production in the production warehouse intended to cover the extraordinary demands. n our case, we do not consider a positive available quantity. If such available quantities are required, they can be considered by entering the required values on the right sides of the constraints. Groups of constraints (10) – (13) ensure the continuity of the inventory status in the warehouse in each time interval. A group of constraints (14) ensures compliance with the free capacity of storage spaces in the production warehouse. A group of constraints (15) will ensure compliance with free loading capacities in the production warehouse. Groups of constraints (16) – (19) define the definition domains of the variables used in the model.

4 Calculation Experiments with the Mathematical Model

Computational experiments with the proposed model were implemented to verify the functionality of the model. To verify the functionality of the model, a model example was used in which it was necessary to produce $Q = 200$ units of production, the planning period was divided into $n = 8$ intervals and the refinement phase was defined by the values of $MINIM = 3$ intervals and $MAXIM = 4$ intervals. The cost of producing one unit is $d = 7{,}5$ monetary units (m.u.), the cost of storing and refining one unit of production is $f = 5$ m.u. and the cost of handling one unit $e = 2{,}5$ m.u.. Vehicles transporting refined products have a capacity of $K_1 = 60$, $K_2 = 100$, $K_3 = 130$ units of production and the costs of accepting them for loading are $N_1 = 60$, $N_2 = 100$, $N_3 = 115$ m.u.. Free available capacities of production, storage spaces and loading in each time interval are shown in Table 2. Some data for a_i and CN_j are missing. The reason is that in these intervals products cannot be produced (the products would not have enough time to refining process) or dispatched (there are no finished products).

Interval *i*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
a_i	130	20	20	20	50	-	-	-
CS_i	200	110	140	100	100	150	150	60
CN_j	-	-	-	4	4	4	1	4

Table 2 Input data

The task is to schedule how many units of production are to be stored in the production warehouse and dispatched from the production warehouse in such a way that the free available production capacity, storage space in the production warehouse, the total loading for production, storage with the refining process and dispatch are kept to a minimum and at the same time there is satisfaction of additional customer requirements. The results are summarized in Table 3. The first column of the table indicates the time intervals. In the second column of the table, the stocked quantities of products are listed. The values in the next columns represent the dispatched quantities in each time interval. The sum of the stocked quantities in each time interval (rows) gives us the values of the stocked quantities x_i. Table 3 also shows compliance with the minimum and maximum values of product stays in the production warehouse. E.g. during the first three intervals there is no dispatch because there are no refined products yet, which corresponds to the $MINIM$ value of 3 intervals.

Interval i	x_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}	z_{i5}	z_{i6}	z_{i7}	z_{i8}
1	110	0	0	0	110	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	20	0	0	0	0	0	20	0	0
4	20	0	0	0	0	0	0	20	0
5	50	0	0	0	0	0	0	0	50
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0

Table 3 Output data

Table 3 shows that free available capacities were observed in production, in the production warehouse, but also at the loading points. The total costs of the process operating in this way is 5, 295 m.u.., of which the production costs are 1, 500 m.u., the costs of storage with the refining process are 3, 000 m.u. and the shipping costs are 795 m.u.. Production of 110 units was dispatched in the fourth interval by vehicle 3. Another dispatch was made for 20 units by vehicle 1 in the fifth and sixth intervals. The last dispatch was carried out in the last interval by vehicle 1 and 50 units of production were loaded. The results of the calculation experiment show that it is more advantageous to start shipping immediately after the product is refined. This phenomenon may seem illogical given the need to use more vehicles, but it is because the model only calculates dispatch and does not consider distribution itself. It would therefore be appropriate to deal with this issue in future research.

5 Conclusion

The presented paper dealt with the issue of determining the volumes of additional production in each time interval of the production process, which also includes the refining phase. The goal was to present a mathematical model enabling the minimization of total costs for the production process, synchronized with the process of storage, refining and dispatching. The functionality of the model was tested on model data for a time period consisting of 8 intervals.

Compared to the original paper [4], the model has been generalized in the sense of increasing the number of vehicle types used for shipping to the customer. The introduction of a heterogeneous vehicle fleet proved to be expedient from the point of view of the needs of logistic practical applications.

The modified model has potential for future research. The potential is in connecting the mentioned process with other important phases of the entire production process (raw material supplies and their subsequent consumption). From the point of view of the needs of logistic practical applications, it is advisable to further expand the model with the possibility of working with heterogeneous production units (glass bottles, PET bottles, cans, barrels), which are widely used mainly in the brewing industry. Some of the mentioned packages also involve reverse logistics or subsequent distribution to different customers, which could expand the model. The last suitable extension seems to be the introduction of uncertainties in the various quantities entering the optimization model.

Acknowledgements

This work was supported by SGS22/125/OHK2/2T/16 Research in the field of computational methods for process optimization in specific postproduction segments of the logistics chains.

References

[1] Aushev, M. (2019). *Model pro optimalizaci distribuce pivovarnických výrobků pro obchodní řetězce*. Praha. Diplomová práce. České Vysoké Učení Technické v Praze.

[2] Georgiadis, G.P., Elekidis, A.P. & Georgiadis, M.C. (2021). Optimal production planning and scheduling in breweries. *Food and Bioproducts Processing*, *125*, 204-221.

[3] Mickein, M., Koch, M. & Haase, K. (2022). A decision support system for brewery production planning at feldschlösschen. *INFORMS Journal on Applied Analytics*, *52*(2), pp.158-172.

[4] Teichmann, D., M. Dorda, & Mocková D. (2019). Model of storage and shipping synchronisation in production warehouses. *Proceedings of the 13th International Conference on Strategic Management and its*

Support by Information Systems (pp. 310-317). Ostrava: VŠB – Technical University of Ostrava, Faculty of Economics.

[5] Tirkeş, G., Çelebi, N. & Güray, C (2021). Developing a multi-stage production planning and scheduling model for a small-size food and beverage company. *Journal Européen des Systèmes Automatisés*, *54*(2), pp.273-281.

[6] Zhang, J. (2021). Application of CPLEX in beer production planning. In *IOP Conference Series: Earth and Environmental Science* (Vol. 831, No. 1, p. 012-034). IOP Publishing.

Regime Switching Behaviour of Selected European Stock Market Returns

Michaela Chocholatá[1]

Abstract. This paper deals with the regime switching behaviour of the selected European stock market returns measured by the DAX index, CAC40 index and STOXX600 index based on the Markov switching (MSW) framework. The paper utilizes the daily data from January, 2018 to January, 2023 and thus enables to capture the behaviour of stock markets during the tranquil as well as turbulent periods. The univariate MSW models were estimated supposing two regimes corresponding to bull and bear market phases characterized by higher returns with lower volatility and lower returns with higher volatility, respectively. The results confirmed the time-varying behaviour of stock returns and corresponding volatility for all analysed returns identifying the turbulent periods corresponding to the global economic slowdown in 2018, as well as to the Covid-19 outbreak in 2020 and to the outbreak of the war in Ukraine in 2022.

Keywords: Markov switching model, regimes, stock market returns

JEL Classification: G15, C22, C58
AMS Classification: 62M05, 62P20, 91B84

1 Introduction

Economic and financial time series occasionally show dramatic breaks in their behaviour associated with various events like economic and political shocks, financial crises, health crises and the war conflicts [6], [9]. Analysing the behaviour of stock market indices thus enables to describe the development of a particular market over time during both the stable periods and times of turbulence [11].

To capture the regime shifts in economic/financial time series, Markov switching (MSW) framework of Hamilton [8] has become very popular. The MSW models introduce a hidden regime variable that follows a Markov chain process to describe the multiple regimes. Parameters of the MSW models are functions of regimes generated by a Markov chain process, see e.g., Dua and Tuteja [7]. Furthermore, the MSW model enables to specify the bull and bear market phases. We recently faced the Covid-19 pandemic which heavily influenced the economic environment including the financial markets all over the world and caused even more substantial downturn than the financial crisis of 2008, see e.g., Ahmed and Sarkodie [1], Bouteska, Sharif and Abedin [4].

Plenty of studies have been published to document and analyse the impacts of this health crisis on the behaviour of stock returns across different countries and areas, e.g., Just and Echaust [10], Dua and Tuteja [7], Reiff [12], Chocholatá [5], [6] and Bouteska, Sharif and Abedin [4]. Various studies have been published to capture the effects of the ongoing war in Ukraine on the behaviour of the world financial markets, e.g., Basdekis et al. [2], Boungou and Yatié [3].

The aim of this paper is to describe the behaviour of the European stock market returns measured by the DAX index, CAC40 index and STOXX600 index analysing the daily data between January, 2018 and January, 2023 with the use of the two-regime Markov switching framework to assess the switching between the "high mean-low volatility" periods (bull regime) and "low mean-high volatility" periods (bear regime) as well as to calculate the transition probabilities between the two regimes.

The rest of the paper is organized as follows. Section 2 is devoted to methodology, section 3 presents data and empirical results of analysis and section 4 concludes.

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, michaela.chocholata@euba.sk.

2 Methodology

Since the recent Covid-19 pandemic and the ongoing war in Ukraine evoked a considerable extent of uncertainty in the global economic growth as well as huge turbulence in the stock market behaviour, this paper uses the MSW model popularized by Hamilton [8] which enables to capture processes driven by heterogeneous states of the world, i.e., the regime-switching behaviour. The basic MSW model distinguishes only two states of the world, i.e. two regimes, corresponding to tranquil "high mean-low volatility" periods (regime 1) and turbulent "low mean-high volatility" periods (regime 2), respectively. The occurrence of the particular regime is determined by an unobservable stochastic process usually denoted as s_t. The regime indicator variable s_t takes on two values – 1 and 2, i.e., $s_t = 1$ indicates that the process is in regime 1 at time *t* and $s_t = 2$ indicates that the process is in regime 2 at time *t*.

In accordance with Hamilton [9] and with regard to the empirical part of the paper, we will suppose that the switching behaviour of the stock returns r_t (continuously compounded returns[2]) could be described by a two-regime MSW model as follows [5], [6], [9]:

$$r_t = c_{s_t} + \emptyset r_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2), \tag{1}$$

where s_t is a regime indicator variable governed by a first-order Markov process, c_{s_t} is the regime dependent intercept and $\emptyset$ denotes the autoregressive parameter. Model (1) enables to capture both the changes in the mean values c_{s_t} as well as changes in the variance $\sigma_{s_t}^2$ between individual regimes.

Transition probabilities for the two-regime Markov chain i.e. probabilities that the regime *i* (at time $t-1$) will be followed by the regime *j* (at time *t*) are specified as follows [6], [7]:

$$P\{s_t = j | s_{t-1} = i\} = p_{ij} \tag{2}$$

Formula (2) specifies the conditional probability which is non-negative and it should hold that:

$$p_{i1} + p_{i2} = 1, \qquad i = 1, 2. \tag{3}$$

3 Data and Empirical Results

The paper analyses the daily closing prices of the selected European stock market indices – DAX (Germany), CAC40 (France) and STOXX600 (as an indicator of the European stock market as a whole) for the period January 1, 2018 and January 31, 2023. The data was retrieved through the R software library "quantmod" from the website finance.yahoo.com [13], the analyses were carried out in R and EViews softwares. As specified in the introduction, the substantial part of the analysis is focused on the corresponding stock returns time series.

Figure 1 illustrates very similar behaviour of the analysed stock market indices and their returns during the whole analysed period, respectively. Although we faced the continued recovery of the global economy in 2017, in 2018 (especially during the last quarter) we witnessed several periods of decline in asset prices and increased volatility which can be attributed e.g., to the global economic slowdown and to the uncertainty about the Brexit issues. Despite the fact that the growth of the global economy slowed down significantly in 2019, stock markets generally recorded the rising trend. However, due to the outbreak and spread of the Covid-19 pandemic, there is a significant collapse of financial markets accompanied by huge volatility during the first quarter of 2020. Another substantial turbulence in the behaviour of the financial markets is connected with the outbreak of the war in Ukraine in February 2022.

To capture the changing behaviour of returns of individual stock market indices during the analysed period, the parameters of the two-regime MSW model (1) were estimated. While for the r_CAC returns the inclusion of delayed observations in (1) was not required, in case of the r_DAX and r_STOXX returns it was necessary to estimate model (1) with AR(1) terms included. Since the mean values and standard deviations were assumed to be regime-specific, parameter $\emptyset$ corresponding to AR(1) term was supposed to be common for both regimes (i.e., regime invariant). The parameter estimates for the two-regime MSW models are included in Table 1. Regarding the statistical significance of the estimated parameters, it can be concluded that the majority of them was statistically significant (with exception of two constants in case of regime 2). The statistically significant negative values of the autoregressive parameter $\emptyset$ indicated the strong cyclical characteristics of the r_DAX and r_STOXX stock

[2] In paper further denoted as returns and denoted as r_DAX, r_CAC and r_STOXX.

return series, respectively. In all three cases[3], regime 1 (bull market) shows positive mean returns and low volatility (0.91%, 0.88% and 0.71%), while regime 2 (bear market) is characterized by negative mean returns and high volatility (2.62%, 2.75% and 1.99%).

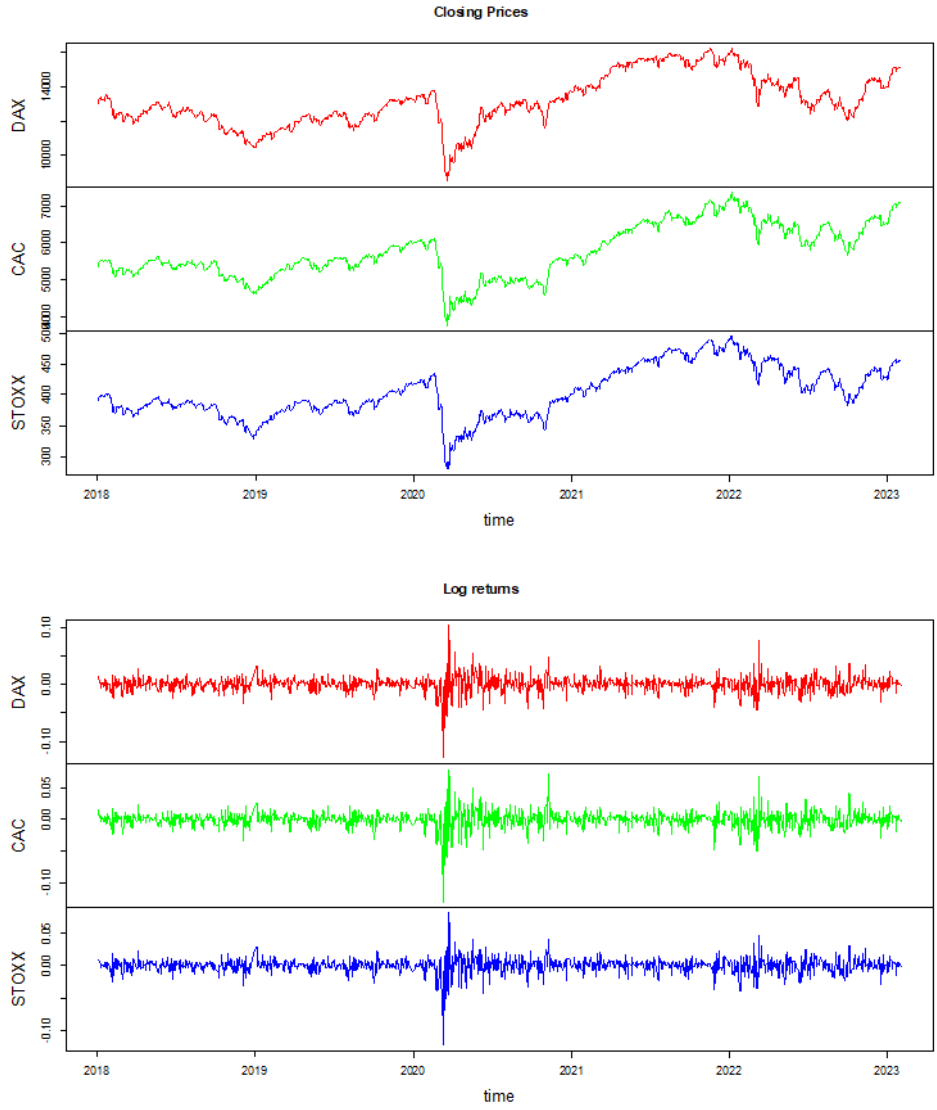


Figure 1 Daily data of stock index prices DAX, CAC40, STOXX600 (upper part) and corresponding stock returns (bottom part)

	Parameter	r_DAX	r_CAC	r_STOXX
Regime 1	c_1	0.0005*	0.0006**	0.0007***
	$\log(\sigma_1)$	-4.6974***	-4.7303***	-4.9533***
Regime 2	c_2	-0.0021	-0.0025	-0.0022*
	$\log(\sigma_2)$	-3.6415***	-3.5950***	-3.9185***
Common	$\emptyset$	-0.0577**	-	-0.0608**

Table 1 Parameter estimates for two-regime Markov switching models. Note: *, ** and *** indicate significance at 10%, 5% and 1% respectively.

Corresponding transition probabilities and expected durations are outlined in Table 2. The probabilities of staying in regime 1 are 0.9806, 0.9862 and 0.9757, respectively, indicating that the regime 1 is much more persistent than

[3] r_DAX, r_CAC and r_STOXX

the regime 2 with corresponding probabilities of 0.8967, 0.9097 and 0.9134, respectively. The transition probabilities from the low volatility regime 1 (bull market) to the high volatility regime 2 (bear market) of 0.0194, 0.0138 and 0.0243, respectively are lower in comparison to transition probabilities (0.1033, 0.0903 and 0.0866, respectively) from regime 2 to regime 1. The expected durations of staying in regime 1, calculated as $\{1/(1-p_{11})\}$, are approximately 51.5405, 72.3329 and 41.1606 days, respectively.

Constant transition probabilities						
	r_DAX		**r_CAC**		**r_STOXX**	
Regime	**1**	**2**	**1**	**2**	**1**	**2**
1	0.9806	0.0194	0.9862	0.0138	0.9757	0.0243
2	0.1033	0.8967	0.0903	0.9097	0.0866	0.9134
Constant expected durations						
Regime	**1**	**2**	**1**	**2**	**1**	**2**
	51.5405	9.6792	72.3329	11.0764	41.1606	11.5428

Table 2 Constant transition probabilities and constant expected durations (two-regime Markov switching models)

The MSW smoothed (ex-post) regime probabilities of being in regime 1 (bull market) at time *t* are shown in Figure 2. Concerning the German and French stock markets, these were broadly in agreement, in terms of identifying the periods of economic turbulence (bear market), however the turmoil periods were identified more often in case of European STOXX600 returns. Furthermore, since the regime switches are random, these cannot be predicted [7], but a high degree of coincidence can be observed between regime switching and periods of economic contractions. Figure 2 illustrates that switching from the tranquil bullish phase (regime 1) to the turbulent bearish phase (regime 2) was the most likely during the short periods in 2018 (mainly in first and fourth quarter), the last quarters in 2019, during the turbulent times in 2020 and many switches between the two regimes are observable from the last quarter of 2021 till the end of analysed period. The described switches to bearish phase are attributable to the economic issues, i.e. to global economic slowdown in 2018, to the ongoing Covid-19 pandemic of 2020 and to the tension and still ongoing war in Ukraine 2022-23.

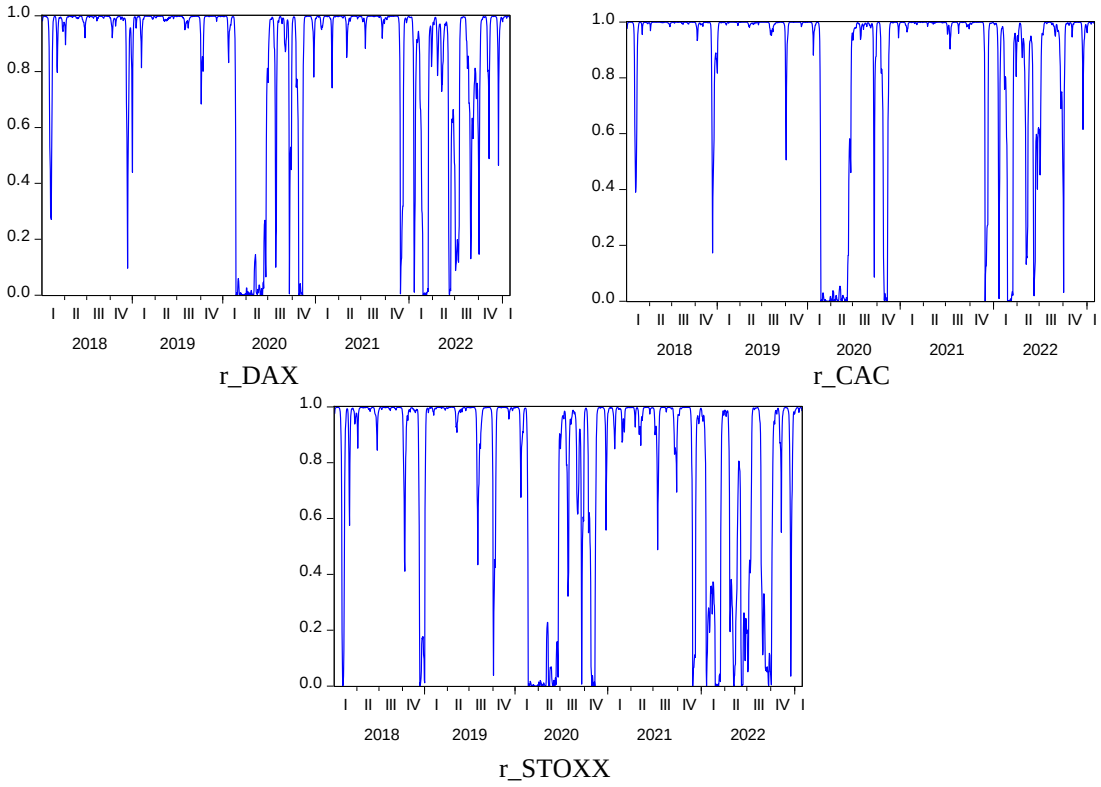


Figure 2 MSW smoothed regime probabilities for the regime 1

4 Conclusion

This paper examined the daily returns of the German DAX, French CAC40 and European STOXX600 stock indices from January 1, 2018 to January 31, 2023. To analyse the regime switching behaviour of these return series, the two-regime MSW models were applied, the included autoregressive parameters (in case of DAX and STOXX600) were not found to be regime dependent. The results showed that the tranquil periods (normal periods, regime 1, bull market) are disrupted by several turbulent periods (crisis periods, regime 2, bear market). The bullish phase (regime 1) was characterized by positive mean returns, whereas in the bearish phase (regime 2) the mean returns were negative. Moreover, the volatility was lower for all the analysed returns during the regime 1 as compared to the higher volatility in regime 2. The transition probabilities of remaining in tranquil periods (regime 1) were higher in comparison to turbulent periods (regime 2) indicating the higher persistence of tranquil periods. The turbulent periods for the analysed returns coincided well with global economic turmoil (global economic slowdown in 2018, the ongoing Covid-19 pandemic of 2020 and the still ongoing war in Ukraine 2022-23).

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0047/23 „The importance of spatial spillover effects in the context of the EU's greener and carbon-free Europe priority. “

References

[1] Ahmed, M.Y. & Sarkodie, S.A. (2021). How COVID-19 pandemic may hamper sustainable economic Development. *Journal of Public Affairs, 21* (4), e2675.

[2] Basdekis, C., Christopoulos, A., Katsampoxakis, I. & Nastas, V. (2022). The Impact of the UkrainianWar on Stock and Energy Markets: A Wavelet Coherence Analysis. *Energies*, *15*, 8174.

[3] Boungou, W. & Yatié, A. (2022). The impact of the Ukraine–Russia war on world stock market returns. *Economics Letters*, *215*, 110516.

[4] Bouteska, A., Sharif, T. & Abedin, M.Z. (2023). COVID-19 and stock returns: Evidence from the Markov switching dependence approach. *Research in International Business and Finance*, *64*, 101882.

[5] Chocholatá, M. (2021). Modelling of PX Stock Returns during Calm and Crisis Periods: A Markov Switching Approach. In R. Hlavatý (Ed.), *International Conference on Mathematical Methods in Economics: Conference Proceedings* (pp. 208–213). Praha: Czech University of Life Sciences Prague.

[6] Chocholatá, M. (2022). Analysis of Stock Returns with Respect to Covid-19 Pandemic and War in Ukraine. In *Quantitative Methods in Economics: Multiple Criteria Decision Making XXI. Proceedings of the International Scientific Conference* (pp. 82–88). Bratislava: Letra Edu.

[7] Dua, T. & Tuteja, D. (2021). Regime Shifts in the Behaviour of International Currency and Equity Markets: A Markov-Switching Analysis. *Journal of Quantitative Economics*, 19 (Suppl 1), S309–S336.

[8] Hamilton, J.D. (1989). A New approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57* (2), 357–384.

[9] Hamilton, J.D. (2010). Regime switching models. In S.N. Durlauf & L.E. Blume (Eds), *Macroeconometrics and Time Series Analysis* (pp. 202–209). The New Palgrave Economics Collection. London: Palgrave Macmillan.

[10] Just, M. & Echaust, K. (2020). Stock market returns, volatility, correlation and liquidity during the COVID-19 crisis: Evidence from the Markov switching approach. *Finance Research Letters*, *37* (2020), 101775.

[11] Reiff, M., Brezina, I. & Pekár, J. (2018). Skrytý Markovov model finančného indexu S&P Europe 350. In *Proceedings from the international conference Trendy v podnikání* (pp.1–8). Plzeň.

[12] Reiff, M. (2021). Vplyv pandémie COVID-19 na akciový index SDAX. In *Proceedings from the conference Využitie kvantitatívnych metód vo vedeckovýskumnej činnosti a v praxi XIV* (pp. 96–100). Bratislava: EKONÓM.

[13] Yahoo. (2023). *Yahoo!finance*. [Online]. Available at: https://finance.yahoo.com/ [cited 2023-02-14].

Measuring and Analyzing the Technical Efficiency of Hockey Players

Lucie Chytilová [1], Jana Hančlová[2]

Abstract. Ice hockey is a trendy sport all over the world. Analysing the efficiency of hockey players is a valuable tool that helps sports managers with player selection, team composition and team performance evaluation. The literature offers only a limited number of scientific studies assessing the efficiency of hockey players or clubs. This research aims to help hockey clubs, managers, and coaches evaluate the efficiency of their players using data envelopment analysis and regression analysis. This research evaluates the technical performance of hockey players using DEA models, ranks the best players, tries to reveal the primary sources of player inefficiency, and uses regression to analyse the change of trend in hockey to evaluate better the future development of the efficiency of hockey players. The models are empirically applied to NHL players over several seasons to see the efficiency gains. The evaluation used in this paper attempts to incorporate greater objectivity into decision-making. It may be an essential step in developing a systematic methodology for evaluating hockey players in the future. However, the results of this first analysis are only visible in the realm of DEA and the possibility of assessing the effectiveness of hockey players. Further developments in regression have yet to identify specific areas of training opportunities.

Keywords: DEA, efficiency, hockey, method of least squares, regression.

JEL classification: C44
AMS classification: 90C05

1 Introduction

Ice hockey is one of the most popular versions of hockey and is spread worldwide, although it is played on an ice surface. And just like in all sports, here, too, it is decided who is the best, the most productive, etc. Ice hockey and efficiency are two topics that can be examined from different points of view. The effectiveness of players, teams, strategies or the use of technology can be increased. All aspects are interrelated and affect these effects on hockey teams and players. Efficiency is essential in achieving success in hockey and is therefore monitored and examined by players, teams and coaches. This article, therefore, aims to help hockey clubs, managers, and coaches evaluate the effectiveness of their players based on player effectiveness. This is the first level of analysis. Thus, various statistics will be used, such as the number of points (goals and assists), time spent on the ice, the score difference during the player's presence (so-called plus/minus) and other factors. These statistics help assess a player's contribution to the team and his ability to contribute to success.

There are many publications on the topic of sports and DEA. For example, a publication by Amin and Sharma [3] focuses on evaluating the performance of English Premier League football clubs from 1998 - 2003. They combine sporting and financial variables to assess how close clubs are to the best practice boundary, analyse how they manage the sport and financial results. In [1], Barros and Leach used DEA for cricket team selection. They proposed a DEA formulation for evaluating cricket players with different abilities using multiple outputs. And for the completeness of the general possibility of use, it should be remembered that, for example, the article [8], where Lozano et al. measure the performance of countries participating in the last five Summer Olympics (using GNP, population and number of medals won). On ice hockey and the DEA, things are already a little worse. However, even here, there are a few publications worth recalling. Kuosmanen, in publication [7], evaluated the effectiveness of NHL ice hockey clubs in the 1996/97 season using data envelope analysis. It includes data on player salaries, league points in the regular season and playoff results in the DEA model. The result is an efficiency score for each team, which tries to help achieve the right team composition and tactical decisions. Hadley et al. in [5] examined the relationship between individual player effectiveness and team effectiveness. It uses traditional radial and additive SBM DEA models. Similar to publication [6] by Jablonský, who analysed the players and statistics of the Canadian-American National Hockey League (NHL) in 2019/2020. His main goal was to combine the relationship between individual player performance and team performance using DEA models. The consulting company PwC (2015) dealt with efficiency in the Czech Republic. Each country's conditions for success at the 2015 Ice Hockey

[1]Czech University of Life Sciences Prague, Kamýcká 129, Praha, Czech Republic, chytilova@pef.czu.cz.

[2]VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, jana.hanclova@vsb.cz.

World Championship in the Czech Republic were analysed. Furthermore, Pelloneová, in publication [11], analysed the technical performance of Czech hockey players using three DEA models. It ranks the best players based on Super Efficiency Score and then looks to uncover the primary sources of player inefficiency, using the 2021/22 Tipsport Extra League seasons. This should then serve for the objectivity of decision-making and the systematic methodology of evaluating hockey players, which we also want to try in this article.

The rest of the article has the following structure: Used Methods, where the fundamental Data Envelopment Analysis (DEA) method and the least squares method are described. Section 3 - Model and Data defines the concrete model and all used variables and gives information about them. Section Results of Analysis - focuses on a close description of the results and description of the future use of this model. Conclusions provide some main results and remarks.

2 Used Methods

The following two subsections describe the methods used in the analysis. It is primarily a data envelopment analysis and then secondarily a basic regression analysis using the method of least squares.

2.1 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric approach. It is widely used for measuring relative efficiency of decision making units (DMUs) with multiple inputs and outputs. Assume, there is a set of T DMUs (DMU$_k$ for $k = 1, ..., T$), let input and output variables data be $X = \{x_{ik}, i = 1, ...R; k = 1, ...T\}$ and $Y = \{y_{jk}, j = 1, ...S; k = 1, ...T\}$, respectively. Also, u_i for $i = 1, ...R$ and v_j for $j = 1, ..., S$ be the weights of the i^{th} input variable and the j^{th} output variable, respectively. Mathematically, the relative efficiency score of DMU$_k$ can be defined as:

$$e_k = \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}}, \text{ for } k = 1, ..., T. \tag{1}$$

Charnes et al. in 1978 in publication [4] have proposed the following CCR model to measure the efficiency score of the under evaluation unit, DMU$_Q$ where $Q \in \{1, ..., T\}$:

$$\begin{aligned} &\max e_Q = \tfrac{\sum_{j=1}^{S} v_j y_{jQ}}{\sum_{i=1}^{R} u_i x_{iQ}}, \\ \text{s.t.}\quad &\textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \le 0, \quad k = 1, ..., T, \\ &u_i \ge 0, \qquad\qquad\qquad\quad i = 1, ..., R, \\ &v_j \ge 0, \qquad\qquad\qquad\quad j = 1, ..., S. \end{aligned} \tag{2}$$

The model (2) is non-linear. It is the model of linear-fractional programming. The model (2) could be transferred by Charmes-Cooper transformation to the standard linear programming problem:

$$\begin{aligned} &\max e_Q = \textstyle\sum_{j=1}^{S} v_j y_{jQ}, \\ \text{s.t.}\quad &\textstyle\sum_{i=1}^{R} u_i x_{iQ} = 1, \\ &\textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \le 0, \quad k = 1, ..., T, \\ &u_i \ge 0, \qquad\qquad\qquad\quad i = 1, ..., R, \\ &v_j \ge 0, \qquad\qquad\qquad\quad j = 1, ..., S, \end{aligned} \tag{3}$$

where $Q \in \{1, ..., T\}$. DMU$_Q$ is CCR-efficient if and only if $e^* = 1$ and if there exists at least one optimal solution $(\mathbf{u}^*, \mathbf{v}^*)$ with $\mathbf{u}^* > \mathbf{0}$ and $\mathbf{v}^* > \mathbf{0}$ for the set $Q \in \{1, ..., T\}$. The inefficient units have a degree of relative efficiency that belongs to interval $[0, 1)$. Note: The model must be solved for each DMU separately.

The model (3) is called a multiplier form of the input-orient-CCR model. However, for computing and data interpretation, it is preferable to work with a model that is dual associated with the model (3). The model is referred to as an envelopment form of the input-oriented CCR model, see [4]. There also exists a multiplier form and envelopment form of the output-oriented CCR model. Both models give the same results, see [4].

Banker et al. in 1984 and their publication [2] have extended the CCR model. The extended model is called the BCC model and considers variable returns to scale assumption. The model has a convex envelope of data which leads to more efficient DMUs. The mathematical model of the dual multiplier form of the input-oriented

BCC model is as:

$$\begin{aligned} & \max e_Q = \textstyle\sum_{j=1}^{S} v_j y_{jQ} - v_0, \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} - v_0 \leq 0, \quad k = 1, ..., T, \\ & u_i \geq 0, \qquad\qquad\qquad\qquad\quad i = 1, ..., R, \\ & v_j \geq 0, \qquad\qquad\qquad\qquad\quad j = 1, ..., S, \\ & v_0 \in (-\infty, \infty), \end{aligned} \tag{4}$$

where v_0 is the dual variable assigned to the convexity condition $\mathbf{e^T}\lambda = \mathbf{1}$ of envelopment form of input-oriented BCC model. Note: The BCC model can be rewritten into the envelopment form or changed into the output orientation.

2.2 Regression Method

Regression analysis is a statistical tool for the investigation of relationships between variables. The usual goals of regression are prediction or control. Regression analysis gives information about the relationship between the chosen explained variable (denoted as Y) by filling in values for the explanatory variables (X) in an equation estimated from data. Regression analysis also offers a measure of the probable accuracy of its predictions. The use of regression analysis for control implies that it could be manipulated with one or more explanatory variables to change the value of the explained variable.

The multiple regression model combines an equation relating the explained variable Y (a.k.a. the dependant variable) to a set of predictors (a.k.a. independent variables) X_1, X_2, ..., X_k with a collection of supporting assumptions. The equation of the model describes:

$$Y_t = \beta_0 + \beta_1 X_{1t} + ... + \beta_k X_{kt} + u_t, \tag{5}$$

where β_i are regression coefficients and u_t are random errors known as residua. Multiple regression models could also be in the form of logarithms or differences.

The optimal multiple regression model describes a data-generating process. The more actual data resemble observations from such an optimal process the more reliable statistical results obtained (such as confidence or prediction intervals).

In addition to the assumed equation (1), the assumptions that complete this multiple regression model are the following:
1. the independent variables X_1, X_2, ..., X_k are non stochastic;
2. the random errors are normally distributed, i.e. u_t $N(0, \sigma^2)$, that is the normality of residues;
3. the mean of random errors is zero, i.e. $E(u_t) = 0$;
4. the variance of random errors is constant, i.e. $\text{Var}(u_t) = \sigma^2$, that is the homoscedasticity;
5. the random errors are uncorrelated, i.e. $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$, that means that model is not autocorrelated;
6. independent variables X_1, X_2, ..., X_k are collinear, i.e. there is no multicollinearity;

3 Model and Data

Data from the NHL site [9] for the Colorado Avalanche, who have won three Stanley Cups (1995–96, 2000–01, 2021–22), will be used for the analysis. Data for the 2002-2003, 2012-2013 and 2022-2023 regular seasons are used for the analysis. This team was chosen as illustrative. At the beginning and end of the observed period, this team was at the top, and in between, it was like when. Plus, for the latest period 2022-2023, data from the website [10] regarding the salaries of hockey players is also used.

Table 1 and 2 show the basic statistical data from all three seasons. Players who played at least 20% of the games in the regular season were selected for each season. We didn't consider the position, but we will deal with this in the future.

In Table 1 and Table 2 are all used variables. The names, definitions and units may be seen below:

- GP - number of played games
- G - number of goals
- A - number of assists
- P - number of points
- plus/min $(+/-)$ - sum of plus and minus points
- PIM - penalty minutes
- PPG - number of power-play goals
- PPP - number of power-play points

season/data															
2002-2003	GP	G	A	P	+/−	PIM	PPG	PPP	SHG	SHP	GWF	OTG	S	S%	FO%
# of obs.	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
max	82	50	77	106	52	88	18	35	2	3	7	2	269	20,49	100
min	32	0	0	1	-12	10	0	0	0	0	0	0	7	0	0
average	63,91	11,36	19,77	31,14	10,86	42,86	3,05	9,05	0,23	0,55	1,86	0,32	105,27	9,66	36,43
st. dev.	17,15	12,17	19,26	29,97	17,20	24,87	4,54	11,52	0,53	0,80	1,86	0,65	75,17	5,36	29,54
2012-2013	GP	G	A	P	+/−	PIM	PPG	PPP	SHG	SHP	GWF	OTG	S	S%	FO%
# of obs.	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
max	48	24	26	43	24	83	24	24	24	24	24	24	132	24	54,65
min	12	0	1	2	-16	0	0	0	0	0	0	0	10	0	0
average	35,21	4,75	8	12,75	-4,58	23,67	0,88	2,54	0,13	0,21	0,58	0,13	58,17	6,96	23,12
st. dev.	10,93	5,26	6,54	11,29	6,18	22,16	1,48	3,79	0,61	0,66	0,83	0,34	34,03	5,06	22,31
2022-2023	GP	G	A	P	+/−	PIM	PPG	PPP	SHG	SHP	GWF	OTG	S	S%	FO%
# of obs.	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
max	82	55	69	111	39	82	13	37	2	3	9	3	366	18	57,14
min	24	0	0	1	-10	2	0	0	0	0	0	0	15	0	0
average	53,46	10,54	17,96	28,50	7,08	24,35	2,46	7,15	0,23	0,54	1,73	0,35	102,04	8,57	22,06
st. dev.	21,39	12,99	18,71	30,18	10,56	18,41	4,06	11,35	0,51	0,90	2,57	0,80	88,47	4,87	22,70

Table 1 Statistical information about used data

2022-2023	Salary
# of observation	23
max	9250000
min	750000
average	3210978,26
st.dev.	2721021,54

Table 2 Statistical information about the saleries of the hokey players in season 2022-2023

- SHG - number of shorthanded goals
- SHP - number of shorthanded points
- GWF - number of game-winning goals
- OTG - number of overtime goals
- S - number of shots
- S% - shooting percentage
- FO% - face-off win percentage
- Salary - salary of NHL player in $

From Tables 1 and 2, it can be read that for each period, we always have over twenty players to assess (22, 24 and 26, respectively). In the first and last observed seasons, the players had the most 82 matches, but in the 2012-2013 season, it was only 48 matches. This also affected the real numbers of goals, assists, points, etc. However, it looks like it should be in proportion at first glance. Some variables are part of others or based on others. Therefore, a good analysis of each model's used variable must be done.

Figure 1 shows the DEA model for hockey players. The selection of variables was as follows. The number of games was not selected, as this may not depend entirely on the player himself. Points were also not chosen because goals and assists are the sum of these two variables. Different numbers of power-play goals and so on are also not considered for DEA, as they are parts of the variables already used. The sum of pluses and minuses was also not selected for DEA, as there are negative or zero values. The same was true for the variable FO%. However, the variables S and S% were used to calculate the input variable - unconverted shots (US) in number, which was calculated as:

$$\mathrm{US} = \mathrm{S} - \frac{\mathrm{S}}{100} \cdot \mathrm{S}\%. \tag{6}$$

This variable should be minimized since it is the unconverted number of goals.

Thus, two variables were chosen as inputs that need to be minimised: US (unconverted shots) and (PIM (penalty minutes). And two variables were selected as outputs that need to be maximised: G (number of goals) and A (number of assists). These variables were chosen to represent the effectiveness of a hockey player in terms of playing efficiency (scoring or recording goals so that he is always as successful as possible and at the same time not injured, for example, when fouling).

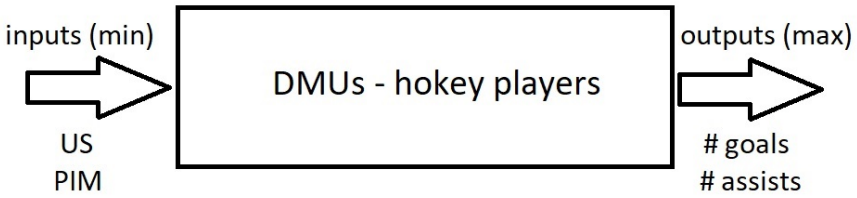


Figure 1 DEA model of the hokey player

The rest of the variables will be used to create the regression model, including the efficiency that will be calculated. And for the 2022 - 2023 season, salary will also be used, but three missing variables need to be replaced; they will be replaced by an average value (average salary in the team). Remember that because some values are zero, the effectiveness was finally calculated for 20, 18 and 21 hockey players. These players are different for all the periods (hokey players changed in the hokey club during these three seasons). The In the future, we will also use DEA models, which may deal with zero or negative values.

For the regression model, we predict that it will meet all the assumptions and, at the same time, we assume the following most likely efficiency relationships (explained variable) and other variables (explanatory variables), see Table 3. Some notes are given in Table 3 as well to see how we had been talking about the relationships, and we made a correlation matrix to understand more about the model. All the calculations had been done in the software

explanatory variable	relationship to explained variable (efficiency)	note
GP	positive *(but)*	little assumption of involvement (the player himself does not affect it)
P	positive	
plus/min	positive	positive values are preferred
PPG	positive	
PPP	positive	
SHG	positive	
SHP	positive	
GWF	positive	
OTG	positive *(but)*	trying to end the game in the regular time
S	positive *(but)*	more shots does not necessarily mean more goals
S%	positive	
FO%	positive	
Salary	positive	money does not necessarily guarantee better results just for the model in season 2022 - 2023

Table 3 Relationships in regression model

STATA and GAMS.

4 Results of Analysis

In Table 4, all efficiencies in all seasons are calculated both according to the CCR model (CCR-ef) and according to the BCC power model (BCC-ef). The table also shows basic statistics. It can be seen that the number of efficient players is always larger in the BCC model. Considering that we did not consider the position of the hockey players, we assess that it is appropriate to choose the BCC model. A model with variable returns from a range that should blur position differences more. In general, the average efficiency is also greater for the BCC models, so it can be seen that the calculations were correct. Based on the above analyses, we decided to use BCC-ef for the regression models.

It should be mentioned that the efficiency in the 2022 - 2023 season is the highest on average (for both models). And in general, it can be seen that the average efficiency is increasing, even though the players played fewer matches (season 2012 - 2013). It can therefore be concluded that the players are improving on average and that their backgrounds or conditions are improving over 20 years. The question is whether this is not also a pressure on the player and whether higher and higher demands can further drain the human organism.

	2002 - 2003			2012 - 2013			2022 - 2023	
	CCR-ef	BCC-ef		CCR-ef	BCC-ef		CCR-ef	BCC-ef
DMU01	1,0000	1,0000	DMU01	1,0000	1,0000	DMU01	0,9534	1,0000
DMU02	0,9385	1,0000	DMU02	0,5446	0,5791	DMU02	1,0000	1,0000
DMU03	0,4364	0,5764	DMU03	1,0000	1,0000	DMU03	1,0000	1,0000
DMU04	0,4198	0,5977	DMU04	0,5571	0,5768	DMU04	0,9109	0,9263
DMU05	0,3491	0,3943	DMU05	0,8917	0,9326	DMU05	0,9916	0,9970
DMU06	0,4988	0,5848	DMU06	0,6392	0,6871	DMU06	0,9226	0,9272
DMU07	1,0000	1,0000	DMU07	0,4746	0,6958	DMU07	1,0000	1,0000
DMU08	1,0000	1,0000	DMU08	0,7854	0,8281	DMU08	0,6026	0,6081
DMU09	0,4541	0,5133	DMU09	0,6360	0,7772	DMU09	0,9713	1,0000
DMU10	0,2030	0,5407	DMU10	0,5345	0,5649	DMU10	0,7554	0,7909
DMU11	0,5007	1,0000	DMU11	0,5220	0,7500	DMU11	0,6465	0,6961
DMU12	0,6574	0,8450	DMU12	0,4313	0,6588	DMU12	1,0000	1,0000
DMU13	0,2861	0,5290	DMU13	0,7868	0,7943	DMU13	0,6085	0,7037
DMU14	0,6500	1,0000	DMU14	0,8307	1,0000	DMU14	1,0000	1,0000
DMU15	0,3946	0,5967	DMU15	1,0000	1,0000	DMU15	0,5697	0,8612
DMU16	0,8143	0,8539	DMU16	0,9716	1,0000	DMU16	0,6656	0,8483
DMU17	0,2809	1,0000	DMU17	0,4286	1,0000	DMU17	0,5692	0,8652
DMU18	0,5193	1,0000	DMU18	0,5370	1,0000	DMU18	0,4661	0,8362
DMU19	0,2345	0,9769				DMU19	0,8539	1,0000
DMU20	0,6400	1,0000				DMU20	0,3509	0,9489
						DMU21	0,3292	1,0000
# eff	3	9	# eff	3	7	# eff	5	9
min	0,2030	0,3943	min	0,4286	0,5649	min	0,3292	0,6081
avergae	0,5639	0,8004	avergae	0,6984	0,8247	avergae	0,7699	0,9052
st. dev.	0,2641	0,2250	st. dev.	0,2096	0,1694	st. dev.	0,2303	0,1203

Table 4 Efficiency score of hokey players
Having the same player designation (DMUxx) across seasons doesn't mean it's the same player in all seasons; on the contrary.

The regression model is therefore defined in such a way that the explained variable Y is BCC-ef, and the explanatory variables are the variables from Table 3. At the beginning of testing, all models were the same for all seasons, except for the 2022 - 2023 season, where the Salary variable is added. However, when testing stationarity, several variables were excluded, and during further testing of the assumptions for the regression model, many variables were excluded, and sometimes the model did not exist. All tests were primarily done at the 5% level of significance, and due to the need for testing, there was an increase to the 10% level of significance. More accurate results are:

- season 2002 - 2003:
  - resulting model: BCC-ef $= \beta_0 + u_t$.
  - properties: the GP variable was not stationary, in the correlation matrix FO% and SHP were significant at the 10% level of significance, but in the linear regression estimation it occurred that neither variable was significant at the 10% level of significance. Plus there are problems in autocorrelation or heteroscedasticity.
  - result: model does not exist!
- season 2012 - 2013:
  - resulting model: BCC-ef $= \beta_0 + \beta_1$SHG $+ u_t$.
  - properties: the GP variable was not stationary, in the correlation matrix SHG was only one and significant at the 10% level of significance. The linear regression estimation showed that it is also significant at the 10% level of significance. There were no problems with another assumptions. The sign in the correlation matrix and the regression was also the same, negative. This does not meet the assumption from Table 3, however, it seems to fit the model.
  - result: Good, but the $R^2 = 10\%$, so it is low.
- sezóna 2022 - 2023:
  - resulting model: BCC-ef $= \beta_0 + \beta_1$FO% $+ u_t$.

- properties: the GP, P, PPG, PPP, GHW, OTG and S variables were not stationary, in the correlation matrix FO% was only one and significant at the 10% level of significance. The linear regression estimation showed that it is also significant at the 10% level of significance. There were no problems with another assumptions. The sign in the correlation matrix and the regression was also the same, negative. This does not meet the assumption from Table 3, however, it seems to fit the model.
- result: Good, but the $R^2 = 10\%$, so it is low.

Nothing extra can be deduced from the results of the regression models. Of the three models, two were more functional and had only one explanatory variable at the 10% significance level. Although the models meet the assumptions for regression models, they are only significant at 10%, which is also not extra. Even though the meaning of the variables is opposite to what was assumed, this area needs to be explored with more and more variables added to avoid the normality problem.

5 Conclusions

Ice hockey is a trendy sport all over the world. Hockey player performance analysis is a valuable tool to assist sports managers with player selection, composition, and evaluation of team performance. This research set out to evaluate the effectiveness of their players using data envelopment analysis and regression analysis to help hockey clubs, managers and coaches. The first part of the analysis, i.e. the DEA models, was successful. They focused on the players' game efficiency by minimising missed shots and penalty minutes and maximising goals and assists. The BCC model was chosen as better, as it did not consider the players' positions. In the case of the second part, the analysis could have been more successful. A regression analysis based on the effectiveness of the BCC model and other possible variables offered by the database (individual subsections of goal types, etc.) and possibly salaries was unsuccessful. No linear regression model would satisfactorily describe the situation and thus provide hockey clubs, managers and coaches with a better overview. So, it is necessary to move this development further.

In the future, it is necessary to expand the number of measurements (more NHL clubs and more years), sort players by position, focus more on player salaries and pay more attention to DEA models - they will use models with non-negative or zero variables. It is also planned to consult with experts in sports training and ice hockey, who can provide a more comprehensive view of the given topic and issue.

References

[1] Amin, G.R. & Sharma, S.K. (2012). Cricket team selection using data envelopment analysis. *European Journal of Sport Science*, 369–376.

[2] Banker, R. D., Charnes, A. & Cooper, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30 (1984), 1078–1092.

[3] Barros, C.P. & Leach, S. (2007). Performance evaluation of the English Premier Football League with data envelopment analysis. *Applied Economics*, 38(12), 1–16.

[4] Charnes, A., Cooper, W. W. & Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (1978), 429–444.

[5] Hadley, L., Poitras, M., Ruggiero, J. & Knowles, S. (2000). Performance evaluation of National Football League teams. *Managerial and Decision Economics*, 21(2), 63–70.

[6] Jablonský, J. (2022). Individual and team efficiency: a case of the National Hockey League. *Central European Journal of Operations Research*, 30(2), 479–494.

[7] Kuosmanen, T. (1998). Efficiency of Hockey Teams in NHL. *Helsinku School of Economics and Business Administration*.

[8] Lozano, S., Villa, G., Guerrero, F. & Cortés, P. (2002). Measuring the performance of nations at the Summer Olympics using data envelopment analysis. *Journal of the Operational Research Society*, 53, 501–511.

[9] NHL. [Online]. Available at: https://www.nhl.com/ [cited 2023-04-20].

[10] NHL Salary Rankings. [Online]. Available at: https://www.spotrac.com/nhl/rankings/cash/ [cited 2023-04-20].

[11] Pelloneová, N. (2022). Measuring the Technical Efficiency of Hockey Players: Empirical Evidence from Czech Hockey Competition. *Studia Kinanthropologica*, 3, 177–186.

[12] PWC - Ice Hockey National Team Performance Model. [Online]. Available at: https://www.pwc.com/gx/en/sports-mega-events/assets/ice-hockey-national-team-performance-model.pdf [cited 2023-04-20].

Application of the Two-Stage DEA Model in SMEs Business

Lucie Chytilová [1], Hana Štverková [2]

Abstract. This study aims to investigate the efficiency of SMEs. First, we introduce a new two-stage DEA model for evaluating firm performance with multi-year variables to measure efficiency based on accounting data in model development. It is primarily concerned with the stable operation of enterprises. The process is divided into two sub-processes: efficiency of human capital (first stage) and efficiency of business (second stage). The outputs of the first stage are the inputs for the second stage. These variables are identified as stocks, investments and economic results of 2020. The external inputs (inputs of the first stage) are also variables from 2020, but the final outputs (outputs of the second stage) are from 2021, which can describe the natural processes in SMEs. The results divide businesses into efficient and inefficient (both overall and in individual phases). Based on these results, the relationships between human capital, business skills and performance are examined. The plan provides recommendations for better functioning and business support in the SME sector.

Keywords: DEA, two-stage, small and medium business.

JEL classification: C44
AMS classification: 90C05

1 Introduction

Small and medium-sized enterprises (SMEs) are the term used to describe a sector of firms of limited size and scope. These businesses typically have fewer employees, lower turnover and a smaller market share than large corporations. The main features of SMEs are:

- size - limited number of employees (generally ranging from a few dozen to a few hundred employees),
- turnover - lower turnover than large corporations,
- independence - firms are owned by private individuals or small groups of entrepreneurs (e.g. family businesses or start-ups, etc.),
- flexibility - they are considered more flexible and able to react faster to market changes with the help of innovation and new trends.

Most countries consider SMEs an important engine of economic growth and job creation. They are often the backbone of the local economy and are a source of innovation and entrepreneurial spirit. There are programmes, financial incentives and advice provided by both governments and non-profit organisations to support the development of SMEs. Therefore, they are still essential today, and therefore a better understanding of these businesses and how they should be effective is important.

There are several ways to measure the effectiveness of SMEs. The most common methods include analysis using financial indicators (turnover, gross margin, cost per employee, profitability and return on investment), operational indicators (capacity, employee productivity, throughput, process time and waste rate), customer satisfaction (customer satisfaction surveys, customer feedback ratings, recommendations, etc.), etc. In general, it is crucial to select appropriate metrics and indicators that match the specific objectives and characteristics of the SME. The combination of different metrics provides a more overall picture of the effectiveness of the business.

In this article, we have chosen the data envelopment analysis method for efficiency measurement and financial indicators, as these are the easiest to obtain. At the same time, several publications already link the given topics, such as Zhou, Ang and Poh in [10] his review article. One of the first connections can be seen in the work of Wu and Liang [9] when they apply DEA to SMEs in China. A more specific focus on the energy industry was then involved in 2015 by Park and Jeong [8]. All this indicates the possibility of use.

We were depending on the complexity of the SMEs business for the model. We came to the opinion that using the two-phase DEA method published in 2012 by Chen, Cook, and Zhu [4] is good. As an aid, we used the article by Li and Feng [7], where the authors also used two phases and two different periods for analysis. However, we primarily used accounting data, which is the beginning of long-term research.

[1] Czech University of Life Sciences Prague. Kamýcká 129. Praha. Czech Republic, chytilova@pef.czu.cz

[2] VŠB - Technical university of Ostrava. Sokolská 33. Ostrava. Czech Republic, hana.stverkova@vsb.cz

The rest of the paper has the following structure: section Two-stage Data Envelopment Analysis provides brief information about the DEA models. Section 3 - Model and Data defines the concrete model and all used variables and gives information about them. Section Results of Analysis - focuses on a close description of the results and description of the future use of this model. The conclusion provides some conclusions and remarks.

2 Two-stage Data Envelopment Analysis

DEA (Data Envelopment Analysis) is a method used to evaluate the efficiency and productivity of units based on their inputs and outputs. Two classical models: the CCR (Charnes et al. [3]) and the BCC (Banker et al. [1]). One variant of DEA is a two-phase DEA, sometimes called network DEA. Färe and Grosskopf [5] were the first to deal with this kind of model. It is an extension where we can combine different inputs and outputs when evaluating the efficiency of units. This makes it possible to model more complex relationships between inputs and outputs and better capture the specifics of units or industries. And it helps evaluate units with a more complex structure of inputs and outputs or requires additional adjustments to measure efficiency more accurately. It can compare units' performance in multi-level systems or industries with diverse characteristics.

There are different types of approaches for the network DEA. In this paper, we use the multiplicative method introduced by Kao and Hwang [6].

Consider the basic input oriented CRS DEA models that estimate the 1st stage, 2nd stage and the overall efficiency for the evaluated unit k_0 independently:

- 1st stage:

$$\begin{aligned} \max \quad & e^1_{k_0} = \frac{\sum_{l=1}^{r} w_l z_{lk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}} \\ \text{s.t.} \quad & \frac{\sum_{l=1}^{r} w_l z_{lk}}{\sum_{i=1}^{m} v_i x_{ik}} \leq 1 \qquad \text{for } k = 1, ..., K, \\ & w_l, v_i \geq \epsilon, \end{aligned} \tag{1}$$

- 2nd stage:

$$\begin{aligned} \max \quad & e^2_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{l=1}^{r} w'_l z_{lk_0}} \\ \text{s.t.} \quad & \frac{\sum_{j=1}^{n} u_j y_{jk}}{\sum_{l=1}^{r} w'_l z_{lk}} \leq 1 \qquad \text{for } k = 1, ..., K, \\ & u_j, w'_i \geq \epsilon, \end{aligned} \tag{2}$$

- overall:

$$\begin{aligned} \max \quad & e_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}} \\ \text{s.t.} \quad & \frac{\sum_{j=1}^{n} u_j y_{jk}}{\sum_{i=1}^{m} v_i x_{ik}} \leq 1 \qquad \text{for } k = 1, ..., K, \\ & u_j, v_i \geq \epsilon, \end{aligned} \tag{3}$$

where k is set of DMUs (DMU$_k$ for $k = 1, ..., K$), let external input, intermediate measures and final output variables data be $X = \{x_{ik}, i = 1, ..., m; k = 1, ..., K\}$, $Z = \{z_{lk}, l = 1, ..., r; k = 1, ..., K\}$ and $Y = \{y_{jk}, j = 1, ..., n; k = 1, ..., K\}$, respectively. Also, v_j for $j = 1, ..., n$ and v_i for $i = 1, ..., m$ be the weights of the i^{th} input variable, w_l for $l = 1, ..., r$ be the weights of the l^{th} intermediate measures on the side of output of the 1st stage, w'_l for $l = 1, ..., r$ be the weights of the l^{th} intermediate measures on the side of input of the second stage, and the j^{th} final output variable, respectively.

To link the efficiency assessments of the two stages, it is universally accepted that the weights associated with the intermediate measures are the same (i.e. $w = w'$), no matter if these measures are considered outputs of the first stage or inputs to the second stage.

In the multiplicative method introduced by Kao and Hwang [6], the overall efficiency and the stage efficiencies of the DMU$_k$ are defined as follows:

$$e^1_{k_0} = \frac{\sum_{l=1}^{r} w_l z_{lk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}}, \quad e^2_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{l=1}^{r} w_l z_{lk_0}}, \quad e^o_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}}, \tag{4}$$

whereas the decomposition model used is

$$e^o_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}} = \frac{\sum_{l=1}^{r} w_l z_{lk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}} \cdot \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{l=1}^{r} w_l z_{lk_0}} = e^1_{k_0} \cdot e^2_{k_0}. \tag{5}$$

i.e. the overall efficiency is the square geometric average of the stage efficiencies.

Given the above definitions, the model below assesses the overall efficiency of the evaluated unit k_o:

$$\begin{aligned} \max \quad & e^o_{k_0} = \frac{\sum_{j=1}^{n} u_j y_{jk_0}}{\sum_{i=1}^{m} v_i x_{ik_0}} \\ \text{s.t.} \quad & \frac{\sum_{l=1}^{r} w_l z_{lk}}{\sum_{i=1}^{m} v_i x_{ik}} \leq 1 \qquad \text{for } k = 1, ..., K, \\ & \frac{\sum_{j=1}^{n} u_j y_{jk}}{\sum_{l=1}^{r} w_l z_{lk}} \leq 1 \qquad \text{for } k = 1, ..., K, \\ & u_j, v_i, w_l \geq \epsilon. \end{aligned} \tag{6}$$

Notice that the constraints $\frac{\sum_{j=1}^{n} u_j y_{jk}}{\sum_{i=1}^{m} v_i x_{ik}} \leq 1$, $k = 1, ..., K$, are redundant and, thus, omitted. Model (6) is a fractional linear program that can be modelled and solved as a linear program by applying the Charnes and Cooper [2] transformation as usual.

3 Model and Data

After much discussion and analysis of the SME market, we decided to use a two-stage DEA model and financial data (most accessible to get) over two time periods. See the model in Figure 3.

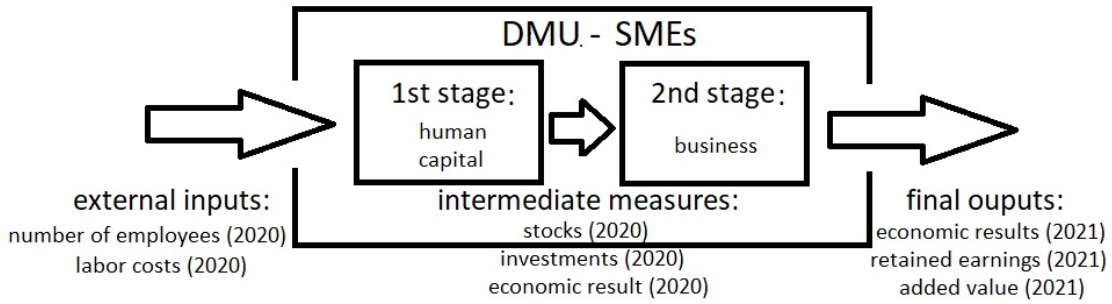


Figure 1 Complete model of the SMEs Business

We divided the model into a first part, which deals with the efficiency of financial management of human capital, and a second part, which deals with the efficiency of the enterprise. As written, SMEs are small regarding number of employees, turnover, independence and flexibility. Therefore, the following variables were important to us:

- external inputs in time period t:
  - number of employees (number) - if the number of employees is appropriate, then the business can function appropriately and save money.
  - labor costs (CZK) - appropriate employee salaries will help both support businesses (people's ideas and growth) and save money for future investments.
- intermediate measures in time period t:
  - stocks (CZK) - on the output side companies want to have as much stocks as possible to fight for the next years, on the other hand from the input side and accounting point of view company want to reduce it as much as possible so that it does not overburden us with storage.
  - investment (CZK) - on the output side company want to increase investment so that companies can grow, innovate and develop. On the input side, company see it as wanting to reduce investment to save money, restructure the business and, at the moment, because of the economic downturn.
  - economic result (CZK) - it is usually the aim of companies to increase economic result, so obviously it is an output, but we have also included it as an input and we wanted to highlight that we want any further economic growth to be greater than the previous period.

- final outputs in time period $t + 1$
  - economic result (CZK) - it is usually the aim of companies to increase economic result for ever.
  - retained earnings - is usually the goal of companies.
  - added value (CZK) - is a goal of many companies, because it can contribute to the financial stability and growth of the company (but not much used today).

For the analysis, 20 firms from the same environment (hospitality) were selected from 2020 and 2021. The specific data can be seen in Table 1. Since some of the data were negative or null, there was still some adjustment with the DEA method, as usual.

id	i1	i2	s1	s2	s3	o1	o2	o3
name of variable	number of employees	labor costs	stocks	invest-ments	economic result	economic result	retained earnings	added value
years	2020	2020	2020	2020	2020	2021	2021	2021
DMU01	37	15789	2674	865	4760	14709	37474	31090
DMU02	48	26037	25962	1200	2578	9664	66387	29322
DMU03	105	63868	60304	11274	27471	71983	89226	63846
DMU04	22	11859	6348	932	2119	2501	9887	14264
DMU05	86	59393	1503	3575	4041	-2614	1203	26918
DMU06	26	10830	8461	368	200	189	12154	13254
DMU07	18	5463	1182	120	570	675	9432	3390
DMU08	2	853	1432	437	186	172	-331	1992
DMU09	6	1986	1567	321	24	11	163	3379
DMU10	19	9764	4563	974	126	-968	8431	4032
DMU11	27	16876	7564	647	2176	2263	39353	25635
DMU12	53	29876	104321	1236	4643	5394	57679	26093
DMU13	8	3809	29	67	440	365	320	7852
DMU14	7	2550	161	543	237	543	2198	3569
DMU15	36	12687	8675	1780	160	432	38654	13975
DMU16	46	15674	12894	1780	246	326	43756	24189
DMU17	19	5567	1260	2100	998	1276	7352	2345
DMU18	48	17317	10903	754	148	243	58763	22393
DMU19	26	10830	8461	357	200	326	12279	13567
DMU20	10	4179	0	1000	-92	-156	16891	16543
average	32.45	16260.35	13413.20	1516.50	2561.55	5366.70	25563.55	17382.40
max	105	63868	104321	11274	27471	71983	89226	63846
min	2	853	0	67	-92	-2614	-331	1992
std.dev	26.58	17327.05	25375.09	2439.00	6076.96	16167.82	26499.66	14804.06

Table 1 SMEs in the Czech Rupublic

4 Results of Analysis

Table 2 shows the result of the all analysis. We can see there:

- e^1 - efficiency of the 1st stage (equation (1)) - *human capital efficiency*,
- e^2 - efficiency of the 2nd stage (equation (2)) - *business efficiency*,
- e - efficiency of the overall model (equation (3)) - *SMEs efficiency*,
- $e^1 \cdot e^2$ - efficiency of multiplication of two stages (equation (5)) - *SMEs multiplication efficiency*,
- e^o - efficiency of the overall multiplicative model (equation (6)) - *SMEs multiplicative efficiency*.

During analysing of the model by stages, it can be seen that 2 SMEs are efficient in the stage of human capital, and 5 SMEs are efficient in the stage of business. According to the model, which would not consider the middle part, only 3 SMEs are efficient. Only one SME is efficient in both parts. It can be seen from these results that being efficient in both human capital and business is very complex and demanding. It can also be seen that the remaining efficient two SMEs have an efficiency in human capital higher than 0.6 and are efficient in business.

So if the SMEs want to be efficient in this case, they should have an efficiency of human capital above average and be for sure efficient in the business. It also means that the fundamental problem of SMEs is wasted human capital. Logically, the results of only multiplying the efficiencies result in only one effective unit being detected. Comparing the overall efficiency based on the multiplicative model, it can be seen at first glance that the results are similar. However, a closer look is more interesting.

	e^1	e^2	e	$e^1 \cdot e^2$	e_o
DMU01	0.8245	0.5454	0.7122	0.4497	0.2928
DMU02	0.8482	0.3959	0.3944	0.3358	0.2057
DMU03	0.9176	1.0000	1.0000	0.9176	0.3881
DMU04	0.4354	0.6606	0.4582	0.2876	0.1393
DMU05	0.1892	1.0000	0.2467	0.1892	0.0537
DMU06	0.3200	0.3707	0.2498	0.1186	0.0905
DMU07	0.5399	0.5085	0.2812	0.2745	0.1717
DMU08	1.0000	1.0000	1.0000	1.0000	0.7628
DMU09	0.5800	0.7154	0.3866	0.4249	0.3087
DMU10	0.2827	0.9314	0.2687	0.2633	0.0691
DMU11	0.8534	0.4575	0.3560	0.3904	0.1998
DMU12	0.6355	1.0000	1.0000	0.6355	0.1445
DMU13	0.6657	0.4270	0.3242	0.2843	0.2173
DMU14	0.5763	0.4672	0.4157	0.2692	0.2902
DMU15	0.7456	0.6827	0.3351	0.5090	0.1792
DMU16	0.6825	0.5407	0.3450	0.3690	0.1640
DMU17	0.4885	1.0000	0.7363	0.4885	0.1830
DMU18	0.8280	0.3164	0.2183	0.2620	0.1990
DMU19	0.3284	0.3659	0.2487	0.1202	0.0927
DMU20	1.0000	0.3006	0.4671	0.3006	0.2913
average	0.6371	0.6343	0.4722	0.3940	0.2222
std.dev	0.2429	0.2628	0.2659	0.3678	0.1541
number of efficient SMEs	2	5	3	1	0

Table 2 Results of efficiency for each stage and overall efficiency

When it came to comparing the three different "overall" effectiveness, it can be said that the "most positive" model (the model with the highest effectiveness) is the *SMEs efficiency* model, (e) compared to others. So it is a model that does not deal with the middle part. This looks nice for analysis, but this model does not give us a closer look at the matter so it can distort the results and certainly does not provide such immediate information.

When the models *SMEs multiplication efficiency* ($e^1 \cdot e^2$) and *SMEs multiplicative efficiency* (e^o) are compared, it can be seen that *SMEs multiplication efficiency* is again more "positive".It is logical again, based on the multiplication. But again, this can be misleading (multiply two low, one low and one high or two high numbers - how much can by synergy and how much is just the reality?). So it looks like a model *SMEs multiplicative efficiency* is for future the best way.

Overall, it is seen that as the best SMEs, all models are defined as the same unit. But if we talk about the worst SMEs, they must explain it similarly. However, SMEs are always at the lower end of efficiency. The only peculiarity is that DMU05 could be better in human resources and more effective in business.

5 Conclusions

This article is an initial longitudinal research in measuring the efficiency of small and medium enterprises and the two-stage method. However, it can be seen that this method is only necessary to choose a suitable potential and model.

Based on the results of the analysis, it can be seen that the fundamental problem is human capital. This part is much harder to balance. And due to today's bleak economic uncertainty, the introduction of artificial intelli-

gence into processes, etc., primarily on the input side, friction surfaces that must be solved can undoubtedly be seen. It is necessary to find room for innovation, but also for human capital and its appropriate evaluation with involvement in these innovations.

In general, when evaluating effectiveness, there is a need to focus more on the use of the model and for whom and for what the analysis can be used. If we would instead show how SMEs are great, the SMEs efficiency model would be suitable. However, if it was a more mindless analysis, then models that work with the efficiency of individual phases would be appropriate. Here it would be best if you thought about the best method. And also see if the process/method suits the hospitality industry.

There is much work to be done here for the future; a better understanding of the environment and an even more thorough analysis of the variables are needed. Use other data models, such as the additive method, the solution of negative variables, different returns to scale, etc.

References

[1] Banker R.D., Charnes A. & Cooper W.W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manage Science*, 30, 1078—1092.

[2] Charnes A. & Cooper W.W. (1962) Programming with linear fractional functional. *Naval Res Logist*, 9, 181–185.

[3] Charnes A, Cooper W.W. & E. Rhodes. (1978) Measuring the efficiency of decision-making units. *European journal of operational research*, 2, 429–444.

[4] Chen, Y., Cook, W.D. & Zhu, J. (2012). Deriving the DEA frontier for two-stage processes. *Omega*, 40(5), 611–618.

[5] Färe R. & Grosskopf S. (1996) Productivity and intermediate products: a frontier approach. *Econ Lett*, 50, 65–70.

[6] Kao C. & Hwang S.N. (2008) Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan. *European journal of operational research*, 185, 418–429.

[7] Li, Y. & Feng, L. (2017). Efficiency evaluation of small and medium-sized enterprises (SMEs) in China based on a novel network DEA model. *Sustainability*, 9(12), 2222.

[8] Park, K. H. & Jeong, B. (2015). The application of data envelopment analysis for evaluating efficiency of small and medium-sized enterprises in the energy industry. *Energy Policy*, 87, 429–438.

[9] Wu, J. & Liang, L. (2012). Efficiency evaluation of small and medium-sized enterprises (SMEs) in China using data envelopment analysis (DEA). *African Journal of Business Management*, 6(4), 1505–1514.

[10] 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.

Best-Worse Method: Comparison with Traditional Approaches

Josef Jablonský[1]

Abstract. Deriving weights of the set of criteria or deriving the priorities of the alternatives with respect to a criterion is an important task in solving multiple criteria decision making problems. Analytic Hierarchy/Network Process (AHP/ANP) is based on a construction of pair-wise comparison matrices (PCM) that contain the expression of preferences of decision makers (DMs). Prioritization methods derive priorities (weights of the criteria, priorities of the alternatives) from the PCM. The most common and popular methods are the eigenvector method, logarithmic least square method and least square method. One of the newer approaches how to derive priorities based on DMs preferences is the Best Worse Method (BWM). The aim of this study is to compare traditional approaches of deriving priorities and the BWM. A special attention is devoted to consistency issues of all considered methods. The illustrative example demonstrates the results of the study.

Keywords: Best-Worse Method, Analytic Hierarchy Process, pairwise comparisons, optimization

JEL Classification: C44
AMS Classification: 91B06

1 Introduction

The analytic hierarchy process (AHP) is a method for solving complex decision problems that was proposed by Thomas Saaty in 1978. Since then, it has become probably the most widely used tool for analyzing decision problems, and the number of applications using this method is certainly in thousands or tens of thousands. AHP is based on structuring decision problems into several subtasks. The entire problem can be expressed as a hierarchical structure similar to that shown in Figure 1. The top level of the hierarchy contains the decision goal. In traditional multicriteria decision-making problems, this is the ranking of alternatives, selection of the best alternative, etc. The next hierarchy level contains usually the criteria that influence the goal. The last level of the hierarchy includes the decision alternatives.

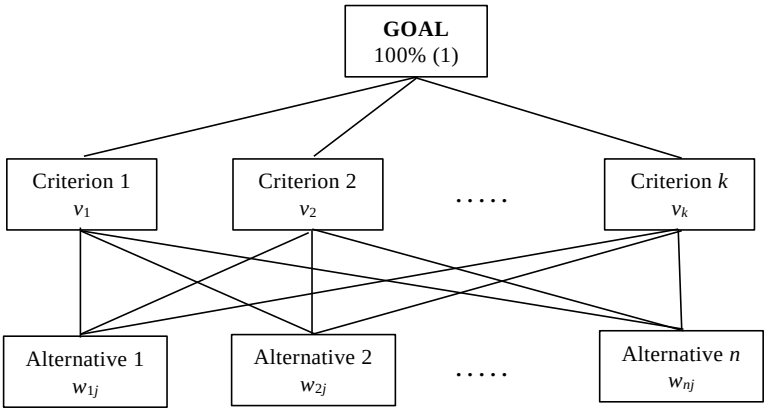


Figure 1 Three-level hierarchy

To derive the priorities of elements (criteria, alternatives) at a level of the hierarchy, it is necessary to judge the relationship of all pairs of these elements. Decision makers compose a matrix of pairwise comparisons whose elements a_{ij}, $i, j = 1, \ldots, n$, where n is the number of elements being compared, express the level of preference of one element over another. The original AHP method - see Saaty [1] - uses 9-point evaluation scale for expression of the relationship where 1 means that the two elements are equally important, and 9 expresses the maximum

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, Praha 3, Czech Republic, e-mail: jablon@vse.cz

(absolute) level of preference of one element over another one. The pairwise comparison matrix **A** is a square reciprocal matrix, i.e. $a_{ij} = 1/a_{ji}$ for all pairs of indices, and $a_{ii} = 1$, $i = 1, \ldots, n$. The crucial problem of the AHP lies in the way how to derive the priorities from pairwise comparison matrices. Many authors discussed this issue in their research. One of the newest approaches is so called Best-Worse Method (BWM) that was proposed in Rezaei [5] and further analyzed in Rezaei [6] and Liang et al. [4]. The BWM became very popular since its publishing. The method itself has many theoretical extensions and real applications.

The aim of this study is to compare the BWM with traditional methods for deriving priorities from pairwise comparison matrices, and answer the question if the BWM really brings any benefit to decision makers. Section 2 provides an overview of prioritization methods in the AHP, Section 3 presents basic features of the BWM, and the further section shows the results of numerical experiments with both approaches. The final section contains discussion of the results and concludes the study.

2 Prioritization Methods in the AHP

Since formulation of basic principles of the AHP, several prioritization methods for deriving priorities from pairwise comparison matrices were proposed. The original Saaty's procedure computes the prioritization vector as the right eigenvector **w** belonging to the largest eigenvalue λ_{max} of the pairwise comparison matrix **A**. This *eigenvector method* consists in solving the following problem:

$$\mathbf{Aw} = \lambda_{max}\mathbf{w} \tag{1}$$

The eigenvector **w** must be normalized, i.e. $\sum_{i=1}^{n} w_i = 1$. Due to the computational problems with solving the problem (1), several approximate prioritization methods were formulated by Saaty and later by other researchers. All of them are based on the minimization of a deviation metric (funkction) between elements of pairwise comparison matrices a_{ij} on one side and ratios of estimated priorities w_i/w_j on the other side.

Least square method (*LSM*) constructs the deviation function as the sum of squares of deviations between elements a_{ij} and ratios w_i/w_j, i.e. the model is as follows:

Minimize
$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij} - \frac{w_i}{w_j}\right)^2 \tag{2}$$

subject to
$$\sum_{i=1}^{n} w_i = 1. \tag{3}$$
$$w_i \geq 0,\ i = 1, \ldots, n.$$

The problem (2)-(3) is a complex non-linear problem with non-unique solutions that are hardly computable. That is why the *LSM* cannot be used for practical purposes. A modification of the *LSM* is the *logarithmic least square method* (*LLSM*) that minimizes the objective function

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\ln a_{ij} - \ln\left(\frac{w_i}{w_j}\right)\right)^2 \tag{4}$$

subject to constraints (3). The solution of problem (3)-(4) can be simply derived as the geometric mean of the elements of each row of matrix **A** that is normalized to the unit sum. That is why this method (originally proposed by Saaty) is often called as *geometric mean method*. Solution of this problem is identical to the eigenvector problem (1) for fully consistent matrices, and it is close to this solution when the consistency measure is on a satisfactory level. More about measures of consistency can be found e.g. in Saaty [7] and Bozoki [1]

Another modification of the *LSM* minimizes the following metric:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij}w_j - v_i\right)^2 \tag{5}$$

The objective function (5) is not linear but it can be transformed into a system of linear equations – see e.g. Bozoki [1] or Gao et al. [2]. Let us denote this method as *modified LSM* (*MLSM*).

Instead of the minimization of the sum of squares it is possible to minimize the sum of positive and negative deviations or to minimize the maximum deviation. In both cases the deviations can be measured either as their absolute values or as relative deviations (in %). The optimization problem for minimization of the sum of relative deviations can be written as follows – let us denote this problem as *RSUM* (*ASUM* for the minimization of the sub of absolute deviations) :

Minimize $$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{d_{ij}^{-}+d_{ij}^{+}}{a_{ij}},$$

subject to $$a_{ij}+d_{ij}^{-}-d_{ij}^{+}=\frac{w_i}{w_j}, \qquad i=1,\ldots,n,\ j=1,\ldots,n, \tag{6}$$

and constraints (3).

A solution that minimizes the maximum relative deviation can be obtained by solving the optimization problem – let us denote this problem as *RMAX* (*AMAX* for absolute maximum deviation):

Minimize $$D,$$

subject to $$a_{ij}+d_{ij}^{-}-d_{ij}^{+}=\frac{w_i}{w_j}, \qquad i=1,\ldots,n,\ j=1,\ldots,n, \tag{7}$$

$$\frac{d_{ij}^{-}+d_{ij}^{+}}{a_{ij}}\le D, \qquad i=1,\ldots,n,\ j=1,\ldots,n,$$

and constraints (3).

The first set of constraints in both problems (6) and (7) is non-linear but their solution can be given quite simply by any non-linear solver, e.g. included in modeling and optimization system LINGO. To avoid non-linearity in models (6) and (7) their simplified version can be formulated and solved. The model for minimization of the sum of deviations is as follows:

Minimize $$\sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}w_j-w_i\right| \tag{8}$$

subject to constraints (3). The model that minimizes maximum deviation is

Minimize $$\max_{i,j}\left|a_{ij}w_j-w_i\right| \tag{9}$$

subject to constraints (3). The models (8) and (9) are not linear but it is possible to re-formulate them as linear models very easily. The solution of models (6) and (7) on one side and models (8) and (9) on the other side is identical only for consistent matrices.

Formulation of all models in this section assumes that all elements of matrix **A** are taken into account either in constraints or the objective function. Due to the reciprocal nature of the pairwise comparison matrix **A** it is questionable whether considering all elements or the elements greater or equal 1 only, i.e. $a_{ij} \ge 1$. All the models presented in this section can be modified accordingly.

3 Best-Worse Method

The BWM was proposed in Rezaei [5] in 2015 and became very popular (almost 3000 citations until now). In fact, the BWM is not a method for solving decision making problems but a method that derives priorities of a set of elements based on their pairwise comparisons. Let us consider, the decision maker intends to derive importance (weights) of the set of *n* criteria (or priorities of the alternatives with respect to a specific criterion). The basic steps of the BWM are as follows:

1. Determining the best (the most important) criterion and the worse (the least important) criterion.
2. Comparison of all criteria with respect to the best criterion on the Saaty's scale (1 to 9) - 9 means that the best criterion is absolutely more important than the specific one. Let us denote the resulting vector of such comparisons (a_{B1}, …, a_{Bn}). Obviously, all elements of this vector are integers between 1 and 9, and at least element equals to 1.
3. Comparison of all criteria with respect to the worse one. The value 9 express that the specific criterion is absolutely more important than the worse one. The resulting vector is denoted (a_{1W}, …, a_{nW}).
4. Based on information included in vectors (a_{B1}, …, a_{Bn}) and (a_{1W}, …, a_{nW}) deriving the preferences (weights) of the criteria by solving optimization problem (10), that minimizes the maximum absolute deviation of the ratio of the weights and elements of the vectors (a_{B1}, …, a_{Bn}) and (a_{1W}, …, a_{nW}). Model (10) is almost identical with model (7) – Jablonský [3].

$$\begin{aligned}
&\text{Minimize} && D, \\
&\text{subject to} && a_{Bj} + d_{Bj}^{-} - d_{Bj}^{+} = \frac{w_B}{w_j}, \quad j = 1,...,n, \\
& && a_{iW} + d_{iW}^{-} - d_{iW}^{+} = \frac{w_i}{w_W}, \quad i = 1,...,n, \\
& && d_{Bj}^{-} + d_{Bj}^{+} \le D, \quad j = 1,...,n, \\
& && d_{iW}^{-} + d_{iW}^{+} \le D, \quad i = 1,...,n, \\
& && \text{and constraints (3).}
\end{aligned} \tag{10}$$

Note that model (10) is identical with model (7) for the modified set of constraints – absolute deviations instead of relative ones. The results of model (10) may not be unique. This is a big problem obtaining different weights of the criteria for the same dataset. Similarly to models (7) and (9), non-linear model (10) can be re-formulated as a linear program (11). Of course the results of model (11) are just approximation of the results of model (10) but they are unique in typical cases. Model (11) is

$$\begin{aligned}
&\text{Minimize} && D, \\
&\text{subject to} && a_{Bj} w_j + d_{Bj}^{-} - d_{Bj}^{+} = w_B, \quad j = 1,...,n, \\
& && a_{iW} w_W + d_{iW}^{-} - d_{iW}^{+} = w_W, \quad i = 1,...,n, \\
& && d_{Bj}^{-} + d_{Bj}^{+} \le D, \quad j = 1,...,n, \\
& && d_{iW}^{-} + d_{iW}^{+} \le D, \quad i = 1,...,n, \\
& && \text{and constraints (3).}
\end{aligned} \tag{11}$$

The methods based on pairwise comparisons as the AHP and BWM must consider the consistency of pairwise comparison matrix (AHP) and pairwise comparisons (BWM). The consistency in the AHP is discussed in the literature in detail. It is measure using consistency ratio as a ratio of the consistency index and random consistency index. Similarly in the BWM, consistency ratio is defined as a ratio of the optimal objective value D* of model (10) and the maximum possible objective value of this model for different maximum pairwise comparison values. They are called consistency indices and their values for n = 3 is 1.00, for n = 5 it is 2.30, and for n = 8 it is 4.47. The values for other maximum values can be found in [5].

The main problem with the definition of consistency in the BWM consists in post-optimization measurement, i.e. to evaluate the consistency it is necessary to solve model (10). In Liang et al. [4] is this way of consistency measurement denoted as output-based consistency. In the contrary, [4] defines input-based consistency that allows evaluating the consistency of pairwise comparisons without solving model (10).

4 A Numerical Illustration

The results of both AHP and BWM methods will be illustrated on a small well-known Saaty's example - estimation of areas of 4 geometric objects – circle, square, rectangle and triangle. They are shown in Figure 2. The importance of the objects is given by their area. The vector of exact results (normalized exact areas of the objects) is (0.490, 0.318, 0.141, 0.051).

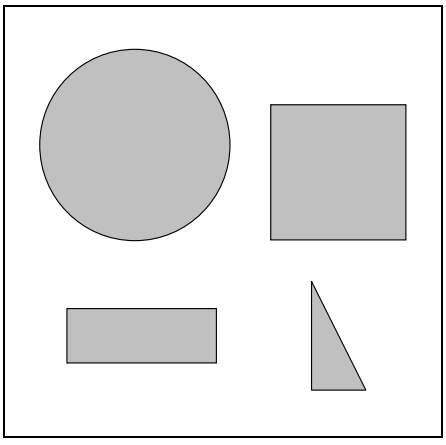


Figure 2 Estimation of the areas of four objects

Two pairwise comparison matrices as inputs for the classical AHP procedure are below. The first matrix fulfills the consistency level very well – the consistency ratio is very low (0.0061). The second matrix has its consistency ratio quite high – 0.0685. But it is still at an acceptable level because the limit for the acceptance is 0.1. Also listed below are two vectors needed for the BWM. Obviously, they are identical with the first row and the last column of the AHP comparison matrix

$$\begin{array}{c} \\ C \\ S \\ R \\ T \end{array}\begin{bmatrix} 1 & 2 & 3 & 8 \\ 1/2 & 1 & 2 & 5 \\ 1/3 & 1/2 & 1 & 2 \\ 1/8 & 1/5 & 1/2 & 1 \end{bmatrix} \qquad \begin{array}{c} \\ C \\ S \\ R \\ T \end{array}\begin{bmatrix} 1 & 5 & 3 & 8 \\ 1/5 & 1 & 2 & 5 \\ 1/3 & 1/2 & 1 & 2 \\ 1/8 & 1/5 & 1/2 & 1 \end{bmatrix}$$

Best to others	(1, 2, 3, 8)	(1, 5, 3, 8)
Others to worse	(8, 5, 2, 1)	(8, 5, 2, 1)
AHP consistency ratio	0.0061 (low)	0.0685 (high)

The results of numerical experiments for both matrices are presented in Tables 1 and 2. Both tables have the same structure. The columns show the results for 8 different models:

- Eigennvalue method (1) – Eigval.
- Logarithmic least square method (4) - LLSM.
- Minimization of the sum of absolute deviations (ASUM) – model (6) with the modified objective function.
- Minimization of the sum of absolute deviations (RSUM) – model (6).
- Minimization of the maximum absolute deviation (AMAX) – model (7) with the modified set of constraints.
- Minimization of the maximum relative deviation (RMAX) – model (7).
- BWM model (10).
- The linearized BWM model (11) - BWML.

The rows in both tables contain:

- The priority vector computed by the method (the sum of elements equals to 1).
- The sum of absolute/relative deviations (*ASUM*/*RSUM*) of decision maker's pairwise comparisons and the ratios of derived priorities.
- The maximum absolute/relative deviation (*AMAX*/*RMAX*).

	Eigval	**LLSM**	**ASUM**	**RSUM**	**AMAX**	**RMAX**	**BWM**	**BWML**
w_1	0.5042	0.5041	0.4848	0.4848	0.4930	0.5001	0.4930	0.5088
w_2	0.2862	0.2864	0.3030	0.3030	0.2958	0.2864	0.2958	0.2719
w_3	0.1456	0.1455	0.1515	0.1515	0.1479	0.1479	0.1479	0.1579
w_4	0.0640	0.0640	0.0606	0.0606	0.0634	0.0656	0.0634	0.0614
ASUM	1.8675	1.8644	1.3458	1.3458	1.7782	2.1823	1.7782	2.3384
RSUM [%]	54.90	54.78	51.67	51.67	53.89	58.70	53.89	71.35
AMAX	0.5281	0.5250	0.5000	0.5000	0.3333	0.6341	0.3333	0.5717
RMAX [%]	15.43	15.49	25.00	25.00	16.67	12.73	16.67	28.58

Table 1 Results for almost fully consistent matrix

The shaded cells show the optimum values for the optimization criteria *ASUM*, *RSUM*, *AMAX*, and *RMAX*. The results for BWM and *AMAX* are identical for consistent matrix and very close in the second case. It is clear because both approaches are based on solving the same non-linear optimization problem. The author of the BWM proposes the linearized version of its model but, as shown in both tables, the results for the original BWM model and the linearized model differ significantly even for the consistent matrix. Another problematic issue in the application of the BWM is the measuring the consistency of the pairwise evaluations. The BWM uses consisten-

cy ratio as a ratio of the optimum objective function D* and the pre-defined consistency index for different maximum values of the pairwise comparison matrix. For our first case, the consistency index is 0.333/4.47 = 0.0745 which is less than the critical value 0.365 (for matrix 4x4). For the second case, the BWM consistency ratio is 1.862/4.47 = 0.417. In the BWM, the second matrix is strongly inconsistent while in the AHP, the consistency level is acceptable.

	Eigval	**LLSM**	**ASUM**	**RSUM**	**AMAX**	**RMAX**	**BWM**	**BWML**
w_1	0.5997	0.5917	0.4848	0.4848	0.5600	0.6039	0.6006	0.5703
w_2	0.2108	0.2126	0.3030	0.3030	0.1784	0.1890	0.1912	0.1406
w_3	0.1320	0.1359	0.1515	0.1515	0.2048	0.1479	0.1473	0.2344
w_4	0.0575	0.0598	0.0606	0.0606	0.0568	0.0592	0.0609	0.0547
ASUM	8.7177	8.1858	4.6458	4.6458	9.7449	8.8467	8.4849	11.8279
***RSUM* [%]**	186.52	177.46	99.67	99.67	243.27	197.00	189.61	300.96
AMAX	2.4296	2.2168	3.4000	3.4000	1.8599	2.2076	1.8621	2.4296
***RMAX* [%]**	51.44	45.13	68.00	68.00	80.30	36.10	37.21	114.26

Table 2 Results for the matrix with borderline level of consistency

5 Conclusions

The BWM is a method for deriving priorities of the set of elements based on pairwise comparisons. In general, it is not a MCDM method even though it may be used, similarly to the AHP, for ranking of alternatives. The idea of structuring the decision making problem as a hierarchy inherently belongs to the AHP and the BWM just uses this idea. The author in [5] discusses some features of the BWM and conclude that the BWM overperforms the AHP in the robustness of the results. He mainly highlights the following features. I would like to argue at least with some of them (original sentences from [5] are in italic):

1. *The BWM is a vector-based method that requires fewer comparisons compared to matrix-based MCDM methods such as the AHP.* That is correct – the AHP requires $n(n-1)$ comparisons while the BWM just $2n-3$. For higher values of n the difference may be significant but one could take into account that well-structured problem has not more than 5 or 6 elements on one level of the hierarchy. For these sizes the advantage of the BWM is not so important.
2. *The priorities derived by the BWM are highly reliable as it provides more consistent comparisons compared to the AHP. While in the AHP, consistency ratio is a measure to check if the comparisons are reliable or not, in the BWM consistency ratio is used to see the level of reliability as the output of BWM is always consistent.* That is not correct at all in my opinion. The BWM measures the consistency level after the priorities are computed by model (10). If the comparisons are not sufficiently consistent, the process must be repeated. In the contrary, the AHP measures the consistency immediately after the matrix is composed. If the matrix is not consistent enough there are available procedures how the consistency may be improved.
3. *Not only can BWM be used to derive the weights independently, it can also be combined with other MCDM methods*. That is not a comparative advantage of the BWM, the same holds for the AHP and other MCDM methods.
4. *While using a comparison matrix, generally speaking we have to deal with integers as well as fractional numbers (e.g. in the AHP we use fractional numbers 1/9, 1/8, …, 1/2, and integer numbers 1, 2, …, 9), in the BWM, only integers are used, making it much easier to use*. That is not a comparative advantage and is just partly correct – the decision makers using the AHP work always with integers (one element is less or more important in certain level).

The claim that the BWM provides better and more robust results than the classical AHP approach is questionable to say the least. There is big problem that the non-linear model (10) need not lead to a unique solution, and its solution is not trivial. The linearized model (11) has always a unique solution but this solution may vary significantly from the solution obtained by model (10). The advantage of the BMW leads mainly in the reduction of necessary pairwise comparisons and user friendliness.

Acknowledgements

The research is supported by Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] Bozóki, S. (2008). Solution of the Least Squares Method problem of pairwise comparison matrices, *Central European Journal of Operations Research,* 16 (3), pp.345-358.
[2] Gao, S., Zhang, Z. & Cao, C. (2009). New methods of estimating weights in AHP. *Proceedings of the ISIP'09*, 2009, Huangshan, pp. 201-204.
[3] Jablonský, J. (2014). Analysis of methods for deriving priorities in pairwise comparisons. In: *Hradec Economic Days 2014*. Hradec Králové: Gaudeamus, pp. 234–240.
[4] Liang, F., Brunelli, M. & Rezaei, J. (2020). Consistency issues in the best worse method: Measurements and thresholds. *Omega*, 96, 102175.
[5] Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53: 49-57.
[6] Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, 64: 126-130.
[7] Saaty, T.L. (1990). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. RWS Publications, Pittsburgh.

Optimizing the Fleet of Transport Ambulances

Ľudmila Jánošíková[1], Peter Jankovič[2]

Abstract. The paper addresses the problem of designing an optimal fleet of transport ambulances. The analysis was ordered by Merea, a.s., which is the second biggest provider of medical transport services in Slovakia. The goal is to propose the number and the locations of the vehicles that could perform planned transports of patients among healthcare facilities. The vehicles are supposed to serve the patients distributed across the whole country. Today, planned transports are accomplished by emergency ambulances. The fully equipped emergency ambulances are necessary in critical cases when the patients need specialised medical care, but they are too expensive for the patients who are not in life-threatening conditions. If a doctor's assistance is not needed during the transport, then the transport can be done using a less equipped, and thus cheaper vehicle. The problem was formulated as a capacitated locations problem and solved using a general integer programming solver. The demand for the model was derived from historical data on planned transports including years 2019 to 2022. Since the mathematical programming model simplifies the reality a lot, we used a detailed computer simulation model to verify whether all required transports could be done with the calculated number and deployment of vehicles. We have found that at minimum 17 vehicles is needed to meet the demand but experiments with a higher number of vehicles were performed as well.

Keywords: ambulance fleet, discrete location, agent-based simulation

JEL Classification: C61, C63
AMS Classification: 90B90, 90C10

1 Introduction

This research optimizes the fleet of vehicles ensuring transports of patients among healthcare facilities. The transports are required when patients cannot get adequate care in the facility where they are hospitalized currently and must be transferred to an appropriate facility. The facilities include not only hospitals and clinics but also rehabilitation centres, dialysis centres, and hospices. The transports are classified into five types: type A refers to the patients who are in critical conditions and need immediate life-saving care; type B concerns an acute injury or disease where the patient needs immediate care in a specialised healthcare facility; types C and D are planned transports and include the patients whose lives are not threatened but they need diagnostics or specialised health care that is not available in the given facility; and finally type E refers to a suspensive transport to another hospital or from a hospital to a home treatment. Currently, the transports of the first four types are ensured by emergency medical service (EMS) ambulances. They are denoted as secondary transports to distinguish them from primary transports elicited by emergency calls. The use of an EMS ambulance is justifiable in case of type A and B transports when a doctor's assistance during the transport is required. In case of C a D types, the presence of a doctor is not needed. Therefore, these transports do not require a fully equipped emergency ambulance, but a cheaper transport ambulance can be used instead.

In this study we design the fleet of transport ambulances. The goal is to decide how many ambulances would be necessary to cover all demand for C a D transports and where these ambulances should be stationed. The service area is the territory of the whole country. We do not consider investment costs of acquisition of vehicles and buildings, where the vehicles and their staff would be housed. Our primary objective is to meet the demand with the minimum-sized fleet. The minor objective is to minimize the waiting time of calls, which is the time that elapses between the dispatching centre receives the call for transport and the arrival of the vehicle at the patient's location.

[1] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, Ludmila.Janosikova@fri.uniza.sk.
[2] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, Peter.Jankovic@fri.uniza.sk.

2 Materials and Methods

The problem of designing an optimal fleet of transport ambulances is a capacitated location problem. Basic location models are described in [2]. Our formulation shares some constraints with the fixed charge capacitated location problem. Parameters of the model are set as follows. The demand has been derived from historical data on planned transports accomplished from January 2019 to October 2022. The spatial and temporal patterns of trips were supplied for us by the Merea company. As regards temporal distribution (Fig. 1 and Fig. 2), the demand is concentrated to weekdays: 87.87% of trips were accomplished from Monday to Friday. Within a day, most transports (72.85%) started between 6 a.m. and 2 p.m. To reduce the model size, we aggregate the demand origins and destinations to central nodes of municipalities, i.e., we suppose that all facilities within a municipality are located at the central node of the municipality. The central node is the road network node that is closest to the centre of the municipality. For every origin – destination (OD) pair of municipalities, the average number of transports per week was calculated. If the origin or destination of the transport were not recorded (for example, if the patient was transported from home), then the transport was excluded. Eventually, the OD matrix contained 93.29% of 148,410 trips that took place in the years 2019 to 2022. It is a sparse matrix, as only 2238 (0.03%) elements have a positive value.

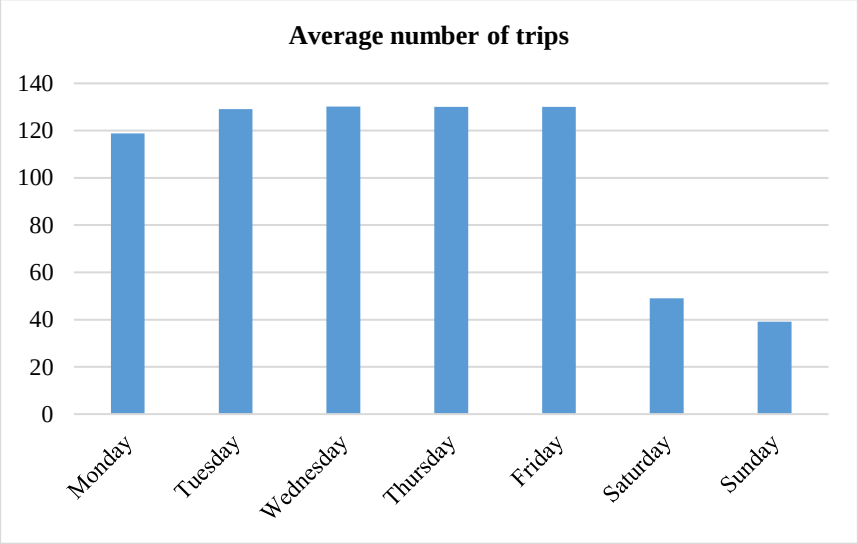


Figure 1 Average number of planned transports per day

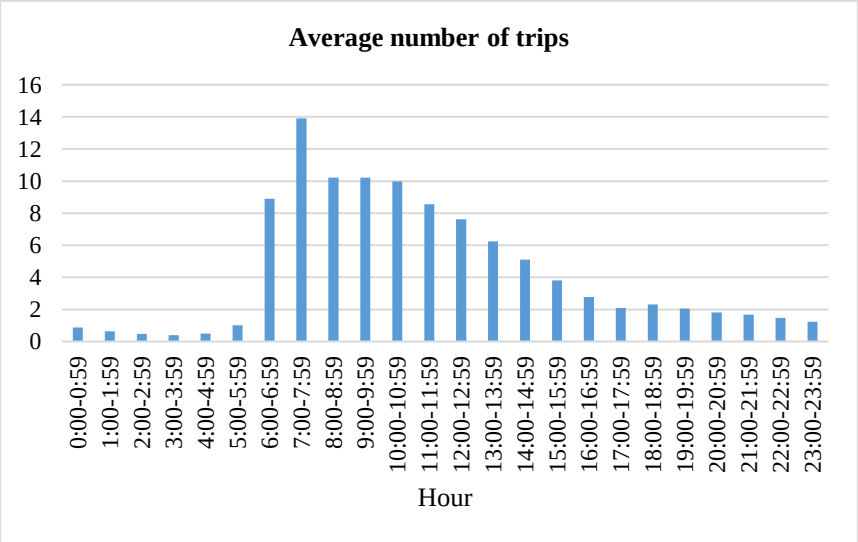


Figure 2 Average number of planned transports per hour

The base station for a transport ambulance can be in every district capital. We did not investigate suitable buildings or lots for a garage, so central nodes of the towns are regarded as candidate locations.

The travel times on the fastest routes between municipalities were calculated using the Dijkstra's algorithm. The digital road network was downloaded from the OpenStreetMap database [6], which is a freely available source of geographical data. The deterministic speeds of vehicles were used. The speed depends on the quality of the road segment and on its location inside or outside a built-up area. Table 1 summarizes the average speeds of vehicles in dependence on the road category; it is adopted from our previous study [5].

Road category	Average speed in rural areas (kilometres per hour)	Average speed in residential areas (kilometres per hour)
Motorway	100	90
Expressway	100	90
Important national road	86	46
National road	67	40
Local road	58	36
Residential road	n.a.	8
Minor road	5	5

Table 1 Average speed of vehicles

In the mathematical formulation of the problem, we will use the following notation:

I	the set of candidate locations; $\|I\| = 108$
J	the set of municipalities; $\|J\| = 2928$
t_{ij}	the shortest travel time from node i to node j (min)
b_{jk}	the weekly number of transports from municipality $j \in J$ to municipality $k \in J$
ts	on-scene time; $ts = 20$ min
th	drop-off time; $th = 15$ min
α	correction factor; $\alpha = 0.8787 * 0.7285$
Q	weekly working time of an ambulance (min); $Q = 5 * 8 * 60 * 0.9329$

The decision variable y_i determines how many transport ambulances will be located at town $i \in I$. The variable x_{ijk} specifies the fraction of demand from municipality $j \in J$ to municipality $k \in J$ transported by vehicles stationed at town $i \in I$.

In the first step, we minimize the number of ambulances needed for the transports demanded on weekdays from 6 a.m. to 2 p.m.:

$$\text{(P1)} \quad \textit{minimize} \quad \sum_{i \in I} y_i \tag{1}$$

$$\textit{subject to} \quad \sum_{i \in I} x_{ijk} = 1 \qquad \textit{for } j \in J, k \in J \tag{2}$$

$$x_{ijk} \leq y_i \qquad \textit{for } i \in I, j \in J, k \in J \tag{3}$$

$$\sum_{j \in J} \sum_{k \in J} \alpha b_{jk} x_{ijk} \left(t_{ij} + ts + t_{jk} + th + t_{ki}\right) \leq Q y_i \qquad \textit{for } i \in I \tag{4}$$

$$x_{ijk} \geq 0 \qquad \textit{for } i \in I, j \in J, k \in J \tag{5}$$

$$y_i \in Z_0^+ \qquad \textit{for } i \in I \tag{6}$$

Constraints (2) ensure that all transports from municipality $j \in J$ to municipality $k \in J$ will be accomplished. Constraints (3) say that transports from municipality $j \in J$ to municipality $k \in J$ cannot be served by vehicles from node i, if there are no vehicles stationed at i. Constraints (4) are capacity constraints. The left-hand side expresses the total time that transports by ambulances from node i take. We suppose that an ambulance travels to municipality j, where the team loads the patient in time ts, then they travel to the destination health facility in municipality k, hand the patient over to the hospital staff in time th, and return to the base. We concentrate on weekdays and peak hours from 6 a.m. to 2 p.m., when the demand is highest. That is why the demand is reduced by the correction factor α. On the other hand, we must account for the demand that was not included in the OD matrix. Therefore, the capacity of vehicles is reduced to make a margin for transports with unknown origins or destinations. Constraint (5) and (6) specify domains of variables.

This model results in the minimum number of ambulances. However, they may be located far away from nodes where the requirements for transport arise. To ensure that patients will be transported as soon as possible, we slightly modify the model. Once the minimum necessary number of ambulances has been determined, we replace the objective function (1) by the optimization criterion (7) that minimizes the travel times of ambulances from their base stations to the origins of the transports. A new constraint (8) is supplemented to constraints (2) – (6) ensuring that the minimum number of ambulances is not exceeded. The resulting model is as follows:

(P2) *minimize* $\sum_{i\in I}\sum_{j\in J}\sum_{k\in J} t_{ij} b_{jk} x_{ijk}$ (7)

subject to (2) – (6)

$\sum_{i\in I} y_i \leq p$ (8)

where p is the objective function value of the optimal solution of the problem P1.

The mathematical model has some drawbacks. The main drawback is the assumption that all trips start at base stations. However, in reality the ambulance can be dispatched to another patient as soon as it becomes available at the destination healthcare facility of the previous trip. Moreover, in the model, travel times as well as service times are deterministic. Although several attempts to model the uncertainty in the ambulance travel time has appeared in the literature, e.g. [1, 3], their incorporation into the mathematical model would result in a stochastic model with high computational complexity. We decided to cope with uncertainty in a different way, namely by using a computer simulation model to evaluate the solution of the mathematical programming model in a more realistic setting.

The computer simulation model still uses deterministic travel times, but they are time dependent. To be concrete, the speed of vehicles reported in Table 1 is reduced during weekdays in morning rush hours (from 6:30 to 9 am), as well as in evening rush hours (from 3 to 6 pm). The on-scene time is modelled by the Erlang probabilistic distribution (shape = 6.44, scale = 3, offset = 1.0) derived from historical data. As soon as the patient has been transferred to the staff in the destination facility, the ambulance is re-dispatched to another trip if there is an unserved call waiting in the queue. The spatial sequence of the trips is optimized, which means that the ambulance is re-dispatched to the closest waiting patient.

The main role of the simulation model is to check whether all demands generated during a day are met. In contrast to the mathematical model, where the demand includes only transports with known origins and destinations, the simulation model considers all transports recorded in the demand matrix. If the origin and/or the destination of the trip are not given, they are generated randomly. The fraction of trips accomplished by 2 p.m. and 6 p.m. is evaluated. These results allow decision makers to decide on the required number of vehicles during day and night.

3 Results and discussion

The mathematical models were implemented in Mosel language and solved using Xpress Optimizer, version 23.01.05. The computational experiments were performed on a PC with Intel® Core™ i5-9300HF CPU, 2.4 GHz and 16 GB RAM. The computing time of the models ranged from 4 to 30 seconds.

To evaluate the fleet performance, we modified the simulation model previously developed for EMS analyses [4, 5]. The model, based on an agent-based architecture and implemented in Java, is a powerful tool designed to simulate and analyse various systems or processes. In its previous version, however, there was a notable absence of transport ambulance modelling. Recognizing this limitation, we have taken the initiative to enhance the model by implementing the necessary agents to account for transport ambulances.

To ensure a comprehensive representation of the system, we have introduced a new generator specifically designed to simulate demand for secondary transport. Starting at 6 am, the model begins generating demands based on carefully analysed real data. These demands capture the actual needs and requirements observed in the analysed dataset, making the simulation more realistic and reflective of real-world scenarios.

The minimum number of ambulances is 17. We calculated 5 different configurations of the fleet by solving the model P2 with parameter p gradually rising from the minimum value of 17 up to 21. Each configuration of the fleet represented one scenario for the simulation experiment. The experiment consisted of 10 replications, each simulating 92 days of system operation. The average results of 10 replications are summarised in Table 2. For each scenario, the following indicators are evaluated:

- fraction of transports finished by 2 p.m., evaluated separately for weekdays and weekends (%),
- fraction of transports finished by 6 p.m., evaluated separately for weekdays and weekends (%),
- average workload of ambulances (%).

Fleet size	Fraction of transports finished by 2 p.m. (%)		Fraction of transports finished by 6 p.m. (%)		Average workload of ambulances (%)
	weekdays	weekend	weekdays	weekend	
17	60.5	83.2	81.9	92.1	52.1
18	63.0	84.2	82.7	91.4	49.1
19	63.1	88.4	84.6	95.1	47.0
20	63.0	94.2	85.3	98.6	44.2
21	63.8	96.3	86.4	99.6	42.1

Table 2 Performance indicators of the fleet of transport ambulances

The results of the simulation experiments indicate that even the minimum-sized fleet of 17 vehicles manages to transport all patients. During weekdays, 82% of transports can be accomplished by 6 p.m. With the increasing fleet size, this percentage raises up to 86%. Due to the rules and habits in operation, health care facilities prefer to admit and discharge patients during the morning shift. That is why we were asked to evaluate an additional indicator, namely the percentage of trips finished by 2 p.m. In the first column of Table 2 we can see that during weekdays this fraction ranges from about 61% to 64%. On Saturday and Sunday, the minimum-sized fleet can accomplish 83% of trips by 2 p.m. and 92% by 6 p.m., while 96% and almost 100% of trips, respectively, can be finished with 21 vehicles. The average workload of ambulances decreases with the increasing fleet size. An average day workload (regardless weekdays or weekend) is 52% for the first scenario with 17 vehicles and it decreases to 42% for the fleet of 21 vehicles.

The simulation model also incorporates several simplifications of the real system. One of them is that ambulances are modelled as working 24 h each day with no breaks for the rest of the crew or technical breaks for cleaning and restocking. Therefore, the calculated values can be regarded as lower estimates of the fleet performance.

4 Conclusions

The combination of mathematical and simulation modelling represents a useful decision support tool. Many simplifications required by the analytical approach can be removed using the simulation model. Computer simulation gives a sufficiently accurate estimate of the performance of the transport services. The proposed configuration and spatial distribution of the fleet must be evaluated from economic point of view. Based on the financial analysis, the decision makers will determine the final number of vehicles and their shifts during a week.

Acknowledgements

This research was supported by the Slovak Research and Development Agency under the project APVV-19-0441 “Allocation of limited resources to public service systems with conflicting quality criteria” and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under the project VEGA 1/0216/21 “Designing of emergency systems with conflicting criteria using tools of artificial intelligence”.

References

[1] Buzna, Ľ. & Czimmermann, P. (2021). On the Modelling of Emergency Ambulance Trips: The Case of the Žilina Region in Slovakia. *Mathematics, 9*, 2165.

[2] Current, J., Daskin, M. & Schilling, D. (2004). Discrete network location models. In Z. Drezner & H. W. Hamacher (Eds.), *Facility Location: Applications and Theory* (pp. 81–118). Berlin: Springer.

[3] Ingolfsson, A., Budge, S. & Erkut, E. (2008). Optimal ambulance location with random delays and travel times. *Health Care Management Science, 11*, 262–274.

[4] Jánošíková, Ľ., Jankovič, P., Kvet, M. & Zajacová, F. (2021). Coverage versus response time objectives in ambulance location. *International Journal of Health Geographics, 20*, 32.

[5] Jánošíková, Ľ., Kvet, M., Jankovič, P. & Gábrišová, L. (2019). An optimization and simulation approach to emergency stations relocation. *Central European Journal of Operations Research, 27*(*3*), 2019, 737–758.

[6] *OpenStreetMap database*. [Online]. Available at: https://www.openstreetmap.org. [cited 2019-04-16].

Queueing Model for Reducing of Waiting Time at Airports

Martin Jasek[1], Ivana Olivkova [2]

Abstract. Reducing the time of the transport process is today a modern trend in all sectors of transport, including air transport. The aim of the work is to focus on the ground side of the air transport process, namely the possibility of reducing the transport time by reducing the waiting time at the airport. In addition to the duration of passenger check-in, we will also focus on calculating the costs incurred in this process. Probably the best method of mathematical modeling of this process is mass service models. The aim of this work is to design a system that would be functional for a various number of incoming passengers, where, depending on the number of passengers registered for the flight, a different number of lines would be used, as needed, so that the processing system would be suitable from both an economic and time (waiting) point of view.

Keywords: air transport, passenger check-in, queueing theory, mathematical modeling, economics, time efficiency

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

As Markovian queueing models have been previously utilized for passenger check-in at airports and have been successful, one of these models will also be employed in this study. The model allows for the computation of the operational characteristics of the queueing system for departing passengers, specifically, a parallel queuing system will be utilized. The passenger security screening process will be studied to simplify the model. [1]

The arrival rate and mean service time follow an exponential probability distribution. The exponential distribution is selected as the calculated parameters are viewed as pessimistic estimates and the exponential distribution represents the most chaos in the system. Passengers enter the system one by one, in the order in which they entered the queue. Therefore, it is a FIFO (First In, First Out) system. The initial input data for the mathematical model of the queueing system were obtained from measurements of the duration of the passenger security check at Ostrava-Mosnov airport. The measurement was taken through experimental measurement of the real check-in process, and the average security screening process time of one passenger (180 s) will be utilized in further calculations as the mean service time. [2] [3]

2 Calculation of Operational Characteristics of Queueing System with Parallel Line Ordering with Queue

In order to design a suitable queuing model, it is necessary to determine the number of passengers that need to be checked in at a given time.

Year	Number of checked-in passengers at Vaclav Havel Airport Prague
2019	17.8 million
2022	10.7 million

Table 1 Number of checked-in passengers at Vaclav Havel Airport Prague [4]

The table above includes the total number of passengers checked-in at Vaclav Havel Airport in Prague. This statistic includes both departing and arriving passengers. However, with the airport's occupancy rate constantly on the rise, it is likely that such values will soon be reached in terms of departing passengers.

[1] VSB-Technical University of Ostrava, Institute of Transport, 17. listopadu 2172/15, 708 00 Ostrava-Poruba, martin.jasek.st@vsb.cz
[2] VSB-Technical University of Ostrava, Institute of Transport, 17. listopadu 2172/15, 708 00 Ostrava-Poruba, ivana.olivkova@vsb.cz, College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava

Calculation of the average number of incoming passengers per hour for the year 2022:

$$\lambda = \frac{\frac{pax}{dcy}}{hpd} = \frac{\frac{10{,}700{,}000}{365}}{19} = 1{,}542.9\ \frac{pax}{h} \Rightarrow \frac{1}{\lambda} = \frac{1}{1{,}542.9}\ h = 2.33\ s \quad (1)$$

Where:
λ – arrival rate (number of passengers arriving per hour)
pax – number of passengers in 2022
dcy – number of days in common year
hpd – number of hours of operation per day (from 5 a.m. to 24 p.m.)
$\frac{1}{\lambda}$ – mean interarrival time

The following input values will be used to determine the operational characteristics of the system:

$$n = 90 \qquad l = 50$$

$$pax = 10{,}700{,}000\ \frac{pax}{y.} \qquad \lambda = 1{,}542.9\ \frac{pax}{h.} \qquad \frac{1}{\mu} = 190\ s$$

Total number of seats in the system:

$$m = n + l = 90 + 50 = 140 \quad (2)$$

Number of all system states:

$$k = m + 1 = 140 + 1 = 141 \quad (3)$$

Where:
n – number of servers
l – number of queueing stands
m – size of the system capacity
k – number of all states
$\frac{1}{\mu}$ – mean service time

Mean number of passengers served by one server:

$$\frac{1}{\mu} = 190\ s = 0.053\ h.$$

$$\mu = \left[\frac{1}{\mu}\right]^{-1} = 18.87\ \frac{pax}{h.}$$

Where:
μ – mean number of passengers served per hour by one server
$\frac{1}{\mu}$ – mean service time

Calculation of arrival rate/service rate:

$$\rho = \frac{\lambda}{n \cdot \mu} = \frac{1{,}542.9}{90 \cdot 18.87} = 0.91 \Rightarrow \rho \neq 1 \Rightarrow \rho \leq 1 \Rightarrow \text{system is stable} \quad (4)$$

For 1,542.9 passengers arriving per one hour, it is necessary to have at least 82 servers in operation, because with fewer servers in operation, the system would become unstable.

$\boldsymbol{\lambda}$	$\boldsymbol{n}$	$\boldsymbol{l}$	$\boldsymbol{\lambda}$	$\boldsymbol{n}$	$\boldsymbol{l}$
1851.48	108	60	925.74	54	30
1697.19	99	55	771.45	45	25
1542.9	90	50	617.16	36	20
1388.61	81	45	462.87	27	15
1234.32	72	40	308.58	18	10
1080.03	63	35	154.29	9	5

Table 2 Number of servers and queueing stands needed to maintain the same traffic intensity for different values of the average number of passengers arriving in the system per hour

As $\boldsymbol{\rho \neq 1}$, the following equations will be used to determine the probabilities of the individual states of the system:

$$P_0 = \left[\sum_{k=0}^{n} \frac{1}{k!} \cdot \left(\frac{\lambda}{\mu}\right)^k + \frac{1}{n!} \cdot \left(\frac{\lambda}{\mu}\right)^n \frac{1-\rho^{m-n+1}}{1-\rho}\right]^{-1}, \text{ applies to } \rho \neq 1 \tag{5}$$

$$P_0 = \left[\begin{array}{c} \frac{1}{0!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^0 + \frac{1}{1!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^1 + \frac{1}{2!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^2 + \\ + \frac{1}{3!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^3 + \cdots + \frac{1}{89!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{89} + \frac{1}{90!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{90} \cdot \frac{1-0.91^{140-90+1}}{1-0.91} \end{array}\right]^{-1} = 2.781 \cdot 10^{-36}$$

$$P_k = \frac{1}{k!} \cdot \left(\frac{\lambda}{\mu}\right)^k \cdot P_0 \quad \text{, applies to } k = 1, \dots, n \tag{6}$$

$$P_1 = \frac{1}{1!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^1 \cdot 2.781 \cdot 10^{-36} = 2.274 \cdot 10^{-34}$$

$$P_2 = \frac{1}{2!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{50} \cdot 2.781 \cdot 10^{-36} = 9.297 \cdot 10^{-33}$$

$$P_{90} = \frac{1}{90!} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{90} \cdot 2.781 \cdot 10^{-36} = 0.025$$

$$P_k = \frac{1}{n! \cdot n^{k-n}} \cdot \left(\frac{\lambda}{\mu}\right)^k \cdot P_0, \text{ applies to } k = n+1, \dots, m \tag{7}$$

$$P_{91} = \frac{1}{90! \cdot 90^{91-90}} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{91} \cdot 2{,}781 \cdot 10^{-36} = 0.023$$

$$P_{92} = \frac{1}{90! \cdot 90^{92-90}} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{92} \cdot 2{,}781 \cdot 10^{-36} = 0.021$$

$$P_{140} = \frac{1}{90! \cdot 90^{140-90}} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{140} \cdot 2.781 \cdot 10^{-36} = 0.209 \cdot 10^{-3}$$

$\boldsymbol{k}$	$\boldsymbol{P_k}$	$\boldsymbol{s}$	$\boldsymbol{l}$
0	$2.781 \cdot 10^{-36}$	0	0
1	$2.274 \cdot 10^{-34}$	1	0
50	$3.889 \cdot 10^{-5}$	50	0
90	0.025	90	0
91	0.023	90	1
140	$0.209 \cdot 10^{-3}$	90	50

Table 3 Certain system states selection

Where:
P_k – probability of k-state
s – number of passengers in the service
l – number of queueing stands

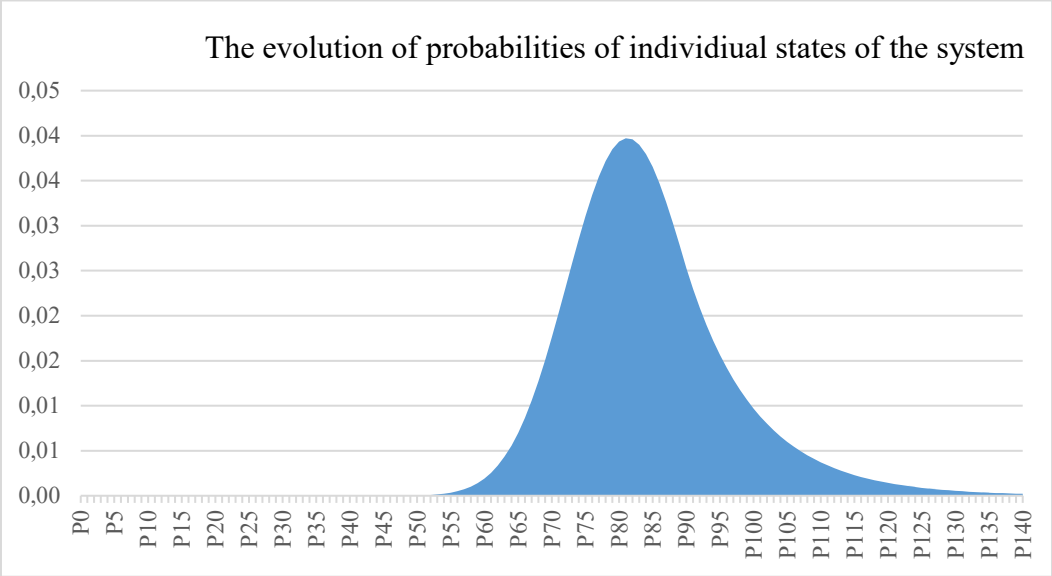


Figure 1 A graph showing the values of probabilities for individual system states

The system starts to be loaded from the moment there are approximately 55 passengers in the system.
Sum of probabilities of all states:

$$\sum_{k=0}^{\infty} P_k = 1 \tag{8}$$

The sum of the probabilities of all states is 100% - this requirement has been met. [5] [6]

Next, the average number of passengers in the system, in the service and in the queue will be calculated:

$$EK = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 + \cdots + 140 \cdot P_{140} \tag{9}$$
$$EK = 0 \cdot 2.781 \cdot 10^{-36} + 1 \cdot 2.274 \cdot 10^{-34} + 2 \cdot 9.297 \cdot 10^{-33} + \cdots + 140 \cdot 0.209 \cdot 10^{-3} = 84.3\ pax$$

$$ES = 0 \cdot P_0 + 1 \cdot P_1 + \cdots + 89 \cdot P_{89} + 90 \cdot (P_{90} + \cdots + P_{139} + P_{140}) \tag{10}$$
$$ES = 0 \cdot 2.781 \cdot 10^{-36} + 1 \cdot 2.274 \cdot 10^{-34} + \cdots + 89 \cdot 0.028 + 90 \cdot (0.025 + \cdots + 0.23 \cdot 10^{-3} + 0.21 \cdot 10^{-3})$$
$$ES = 81.7\ pax$$

$$EL = EK - ES = 84.3 - 81.7 = 2.6\ pax \tag{11}$$

Where:
EK – average number of passengers in the system
ES – average number of passengers being served
EL – average queue length

Probability of rejecting a passenger and average utilization of each line:

$$P_{odm} = \frac{1}{n! \cdot n^{m-n}} \cdot \left(\frac{\lambda}{\mu}\right)^m \cdot P_0 \tag{12}$$

$$P_{odm} = \frac{1}{90! \cdot 90^{140-90}} \cdot \left(\frac{1{,}542.9}{18.87}\right)^{140} \cdot 2.781 \cdot 10^{-36} = 0.209 \cdot 10^{-3}$$

$$\lambda \cdot P_{odm} = 1{,}542.9 \cdot 0.209 \cdot 10^{-3} = 0.32\ pax \cdot h.^{-1} \tag{13}$$

$$\kappa = \frac{ES}{n} = \frac{81{,}7}{90} = 0.91 \tag{14}$$

Where:
P_{odm} – probability that a passenger is rejected
$\lambda \cdot P_{odm}$ – mean number of rejected passenger per hour
κ – utilization factor of a server

Finally, calculation of the passenger's mean waiting time for service and the passenger's mean stay in the system:

$$EW = \frac{\mathrm{EL}}{\lambda \cdot (1 - P_{odm})} = \frac{2.6}{1{,}542.9 \cdot (1 - 0.209 \cdot 10^{-3})} = 1.685 \cdot 10^{-3}\ hod. = 6\ s \tag{15}$$

$$EZ = \frac{1}{\mu} + EW = \frac{1}{18.87} + 1.685 \cdot 10^{-3} = 0.05\ hod = 196\ s \tag{16}$$

Where:
EW - average waiting time in a queue
EZ - average time in the system

Calculation of operational costs:

The Vaclav Havel Airport in Prague charges 749 CZK per person for the use of the airport by passengers. The hourly operation cost of the counter with year-round rental is 134 CZK.

$$\lambda = 1{,}542.9\ pax./h. \qquad P_{odm} = 0.209 \cdot 10^{-3}$$

$$n_z = p_{plc} + \frac{n \cdot p_{ph}}{\lambda - \lambda \cdot P_{odm}} = 749 + \frac{90 \cdot 134}{1{,}542.9 - 1{,}542.9 \cdot 0.209 \cdot 10^{-3}} = 756.82\ CZK \tag{17}$$

Where:
n_z – the costs incurred by passengers
p_{plc} – passenger fee for using the airport
p_{ph} – the price of hourly counter operation during yearly rental
λ - arrival rate (in this case, it is also the mean number of passengers served by servers per hour, since no passenger is rejected in one hour)

Average hourly cost that passengers would have to pay together:

$$n_c = n_z \cdot \lambda \cdot (1 - P_{odm}) = 756.82 \cdot 1{,}542.9 \cdot (1 - 0.209 \cdot 10^{-3}) = 1{,}167{,}454\ CZK \tag{18}$$

n	9	27	45	72	90
λ	154.29	462.87	771.45	1,234.32	1,542.9
κ	0.85	0.89	0.9	0.9	0.91
EL	1.14	2.57	3.05	2.9	2.6
EK	8.76	26.68	43.57	67.95	84.3
ES	7.62	24.1	40.52	65.05	81.7
EW	28.5	20.25	14.292	8.46	6.066
EZ	219.28	211.03	205.07	199.24	196.84
n_z	757.37	756.92	756.85	756.82	756.82
n_c	109,070	345,800	581,498	929,945	1,167,454

Table 4 System operational characteristics based on the number of lines in operation

The table above shows the calculated operational characteristics of the queuing system for five different numbers of servers in operation while maintaining the same arrival rate/service rate. The number of servers and queueing stands here is growing in direct proportion. It is worth noting that although the hourly cost of running the entire system increases in direct proportion to the number of lines in operation, the unit cost per passenger changes only slightly.

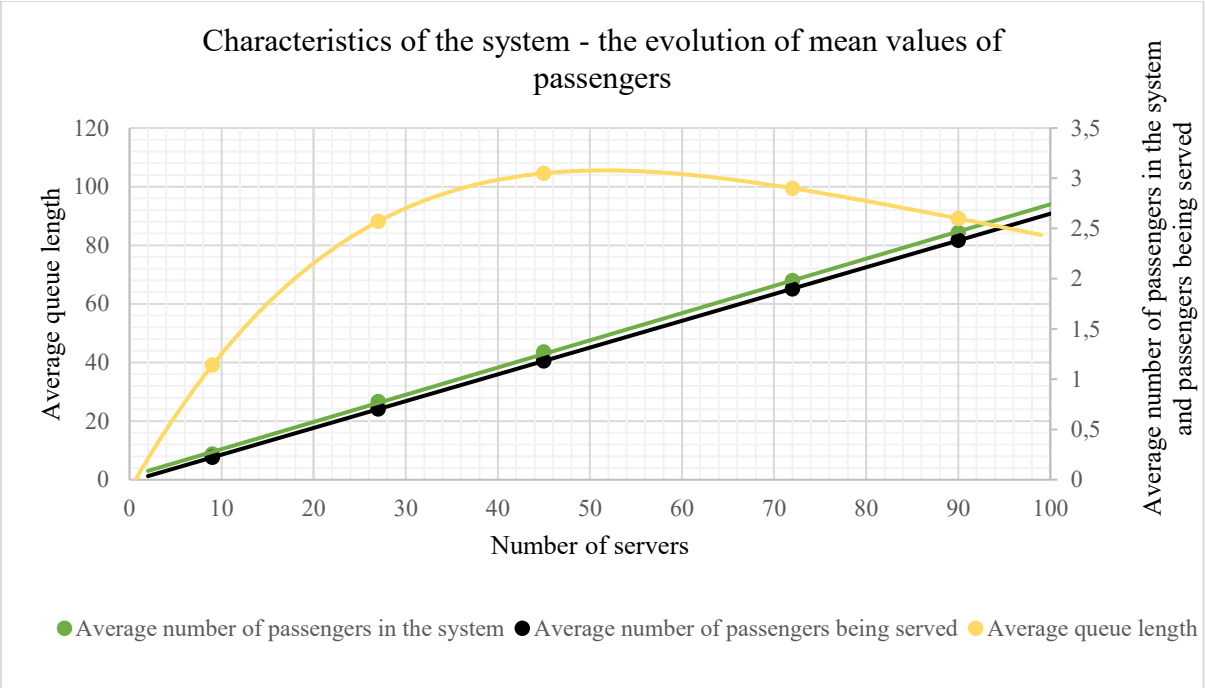


Figure 2 A graph showing the evolution of values EL, EK and ES

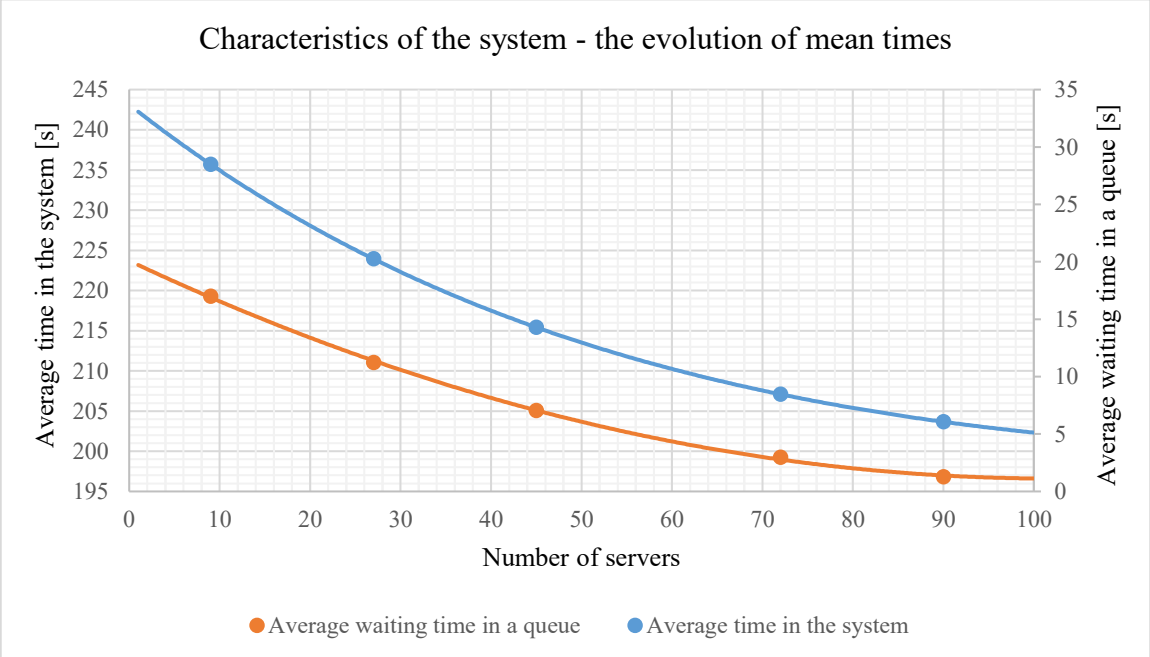


Figure 3 A graph showing the evolution of mean times EW and EZ

The evolutions of selected operational characteristics of the system are shown in the graphs above. The trends of these characteristics were always obtained by fitting five points with a mathematical curve (see Table 4), where the operational characteristics of each model were calculated using the same procedure as the above example for n = 90 and l =50. [4] [7]

3 Discussion and Conclusion

If we wanted to maintain a low probability of passenger rejection in the system with the same incoming passenger intensity and fewer servers in operation, passengers would have to wait longer in the queue. However, since the number of passengers arriving in the system fluctuates throughout the day, it would be beneficial

to determine the hourly changes in the number of passengers and open the necessary number of servers accordingly.
The presented figures and tables can assist in estimating some of the operational characteristics of the system at varying numbers of passengers arriving in the system. Additionally, the attached tables allow for estimating hourly costs, both for individual passengers and for the operation of the entire system. This queueing model could be applied to airports with varying traffic intensities. [8]

Acknowledgements

This work was supported by project SP2023/087 Applied research, experimental development and innovation in transport and logistics.

References

[1] Ademoh, N. A. & Nneka, E. (2014). Queuing Modelling of Air Transport Passengers of Nnamdi Azikiwe International Airport Abuja, Nigeria Using Multi Server Approach. *Middle-East Journal of Scientific Research*, p.13 [Online]. Available at: https://www.idosi.org/mejsr/mejsr21(12)14/17.pdf [cited 2023-01-10].

[2] Simaiakis, I. & Balakrishnan, H. (2015). A Queuing Model of the Airport Departure Process. *Transportation Science,* 50(1), 94-109. https://doi.org/10.1287/trsc.2015.0603

[3] Sodek, L. (2020). *Simulace bezpecnostni kontroly cestujicich na letisti*. Bachelor thesis, VSB – Technical University of Ostrava. Thesis supervisor Michal Dorda. Ostrava. [Online]. Available at: https://dspace.vsb.cz/bitstream/handle/10084/141348/SOD0016_FS_B3712_3708R037_2020.pdf?sequence=1&isAllowed=y [cited 2023-02-21].

[4] Anon. (2022). Zprava o udrzitelnem rozvoji skupiny Letiste Praha: *Zprava skupiny Letiste Praha za rok 2021*, Praha. [Online]. Available at: https://www.prg.aero/profil-csr-zprava [cited 2023-02-23].

[5] Dorda, M. (2011). *Teorie hromadne obsluhy*, Ostrava: VSB - Technical University of Ostrava. [Online]. Available at: http://homel.vsb.cz/~dor028/Hromadna_obsluha.pdf [cited 2023-02-17].

[6] Voracova, S. (2020). *Teorie hromadne obsluhy (Queuing Theory)*, Praha: Department of Applied Mathematics, CVUT. [Online]. Available at: https://www.fd.cvut.cz/department/k611/pedagog/K611THO.html [cited 2023-02-26].

[7] Drever, J. (2020). Data modelling to reduce queuing times through airports. *Airport World's magazine*. [Online]. Available at: https://www.snclavalin.com/en/beyond-engineering/data-modelling-to-reduce-queuing-times-through-airports [cited 2023-02-16].

[8] Mehri, H. Djemel, T. & Kammoun, H. (2008). Solving of waiting lines models in the airport using queueing theory model and linear programming the practice case. *HAL*, p.26. [Online]. Available at: https://hal.science/hal-00263072/document [cited 2023-02-10].

Optimizing the Selection of a Portfolio of Transport Infrastructure Investment Projects Including Elements of Uncertainty Modeled Using Fuzzy Logic

Karel Ječmen [1], Daniel Pilát [2], Dušan Teichmann[3], Denisa Mocková[4], Olga Mertlová[5]

Abstract. Along with the development of the society, emphasis is placed on the modernization and development of quality transport infrastructure. Investing in infrastructure is implemented using investment projects, which are financed in the vast majority from public sources. The amount of the available budget is usually lower than the amount of funds needed to implement all the projects being prepared, and therefore it is necessary to select a portfolio of investment projects for implementation. During the selection process, indicators representing the quantified societal benefits resulting from the implementation of projects, as well as their financial demands and the available budget, are assessed. However, the selection is carried out in the initial phase of the preparation of investment projects, and due to the preparation period of several years, the implementation takes place in a few years. The development of the economic situation has been very turbulent and fluctuating in recent years, while the situation at the time of project implementation often does not correspond to the expected market development in the initial phase of project preparation. As a result, the amount of financial indicators and the estimated available budget at the time of implementation are significantly different than what was assumed during the selection. This can have a negative effect on the output of the optimization of the selection of the project portfolio from the point of view of fulfilling the overall societal benefit.
The aim of the article is to present an approach to optimizing the selection of a portfolio of transport infrastructure investment projects for implementation, including elements of uncertainty in the area of the available budget and the estimated amount of funds needed to implement individual investment projects. Uncertainty is modeled using fuzzy logic, and the goal of optimization is to maximize the cumulative value of social benefits represented by monitored indicators.

Keywords: linear programming, fuzzy, transport infrastructure, optimization

JEL Classification: C61
AMS Classification: 90C47

1 Introduction

Transport infrastructure development and modernization play an important role in enhancing economic growth and societal well-being [4]. Investment in infrastructure is usually financed using public funds, and selecting projects portfolio to implement is often a challenging task for decision-makers [6]. The selection process involves assessing various indicators that represent the quantified societal benefits resulting from the implementation of projects, as well as their financial demands and the available budget as described in [3].

However, the selection is carried out in the initial phase of the preparation of investment projects, and due to the preparation period of several years, the implementation takes place in a few years. The development of the economic situation has been very dynamic and fluctuating in recent years, while the situation at the time of project

[1] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, jecmekar@fd.cvut.cz.
[2] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, pilatdan@fd.cvut.cz.
[3] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, teichdus@fd.cvut.cz.
[4] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, mockova@fd.cvut.cz.
[5] Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport, pokorol1@fd.cvut.cz.

implementation often does not correspond to the expected market development in the initial phase of project preparation. As a result, the amount of financial indicators and the estimated available budget at the time of implementation could be significantly different than what was assumed during the selection. This can have a negative effect on the output of the optimization of the selection of the project portfolio from the point of view of fulfilling the overall societal benefit.

To address this challenge, this article proposes an approach to optimize the selection of a portfolio of transport infrastructure investment projects for implementation, including elements of uncertainty in the area of the available budget and the estimated amount of funds needed to implement individual investment projects. The proposed approach employs fuzzy logic, which can handle the ambiguity and vagueness that are inherent in real world decision-making problems [8]. The goal of the optimization is to maximize the cumulative value of social benefits represented by monitored indicators.

The article presents a valuable contribution to the field of infrastructure investment, as it provides a framework for decision-makers to make more informed choices when selecting which projects to prioritize. By incorporating uncertainty into the optimization process, the article offers a more realistic and effective approach to infrastructure project selection. Fuzzy objectives and constraints can be defined as fuzzy sets in the set of alternatives. Fuzzy decisions can therefore be viewed as the intersection of given objectives and constraints. The maximization is defined as the point in the space of alternatives at which the membership function of the fuzzy decision reaches its maximum value [1].

Fuzzy logic has found applications in various fields, including control systems, artificial intelligence, decision analysis, pattern recognition, and optimization. It provides a more flexible and human-like approach to modelling and reasoning in situations where uncertainty, ambiguity, or imprecision are present. By capturing and representing uncertainty in a quantitative manner, fuzzy logic allows for more robust and adaptive systems that can handle real-world complexity. The proposed method has the potential to improve the overall societal benefit of infrastructure investment projects, and can be applied to other fields facing similar challenges, such as environmental management [9] or renewable energy [2].

In summary, the proposed approach provides a useful tool for decision-makers to optimize the selection of a portfolio of transport infrastructure investment projects, taking into account the uncertainty that often characterizes such decisions. By doing so, the proposed method can contribute to more effective and efficient allocation of resources and lead to a more sustainable development of transport infrastructure.

2 Problem Formulation

Consider a set of investment projects I intended for implementation on transport infrastructure and a set of indicators J. Indicators can be understood as quantified society-wide benefits, which are used to evaluate the development of transport infrastructure. Every project $i \in I$ contributes to the fulfilment of the indicator $j \in J$ with value a_{ij} and the total amount of funds available for investment in transport infrastructure N drains costs in the financial amount n_i.

We suppose that at the moment when the selection of projects for implementation is decided, the monetary parameters of the task cannot be precisely determined. Exact discrete values of the budget and costs for the implementation of individual projects are expressed by triangular fuzzy numbers corresponding to the expected possibilities in the implementation horizon based on an expert estimate of the contracting authority. The aim of the task is to determine which projects will be included for implementation so as to maximize the cumulative value of indicators of implemented projects and at the same time to take into account the uncertainty occurring in the task.

Recapitulation of input parameters of crisp optimization model:

I set of projects to be implemented

J set of monitored indicators

n_i cost of project $i \in I$ implementation

a_{ij} project $i \in I$ contribution to the indicator $j \in J$ fulfilment

N total amount of available funds

Variables used in the model:

x_i binary variable modelling project $i \in I$ implementation

When the optimization calculation is completed with value $x_i = 1$, the project will be selected for implementation. When the optimization calculation is completed with $x_i = 0$, the project will not be selected for implementation.

A fuzzy linear model is used to solve the problem.

3 Mathematical Model

In this chapter, the creation of a fuzzy mathematical model is explained. First, a crisp mathematical model is created, which is then transformed into a fuzzy model.

The difference between crisp and fuzzy mathematical models lies in how they handle uncertainty. Crisp models assume precise values, while fuzzy models allow for the representation of imprecise or uncertain information using fuzzy logic. Fuzzy models provide a more flexible and realistic approach to modelling uncertain or vague systems, whereas crisp models are more suitable for situations where certainty is assumed or uncertainty can be ignored. [5]

Crisp Mathematical Model

The crisp mathematical model has the following form:

$$f(x) = \sum_{i \in I} \sum_{j \in J} a_{ij} x_i \rightarrow max \tag{1}$$

subject to:

$$\sum_{i \in I} n_i x_i \leq N \tag{2}$$

$$x_i \in \{0,1\} \qquad \text{for } i \in I \tag{3}$$

The objective function (1) expresses the value of the optimization criterion, which is the total fulfilment of the monitored indicators. The constraint (2) ensures that the limit of drawing of available funds for the implementation of projects is not exceeded. The group of constraints (3) specifies the definition scopes of variables in the model.

Fuzzy Parameters

Consider that the monetary parameters, in this case the total amount of available funds N and project $i \in I$ implementation costs n_i are fraught with uncertainty, and their discrete value cannot be accurately determined. However, it is possible to determine the limit values of these parameters by expert estimation. In general, both fuzzy parameters can be expressed as $\tilde{n}_i = [n_{i1}; n_{i2}; n_{i3}]$ and $\tilde{N} = [N_1; N_2; N_3]$.

These intervals express the belonging of values in given intervals to the fuzzy parameter; see Figure 1. Minimal project implementation costs n_{i1} are specified as a value corresponding to the project $i \in I$ costs discounted using a moderate inflation scenario, median n_{i2} represents budgetary costs corresponding to the medium inflation scenario, and the extreme value n_{i3} is determined by the financial value of the project, which is defined as the highest possible cost of project implementation while maintaining the economic efficiency of the investment. The values of the budget interval are determined similarly based on the assumptions of the fulfilment of transport policy. The constraint (2) can be written in the form

$$\tilde{\boldsymbol{n}}(x) \leq \tilde{\boldsymbol{N}},$$

where the left side of the constraint $\tilde{\boldsymbol{n}}$ and the right side $\tilde{\boldsymbol{N}}$ must be understood as an expression of one fuzzy number. According to fuzzy logic, these values can be expressed as triangles, where

$$n_1(x) = \sum_{i \in I} x_i n_{i1}, n_2(x) = \sum_{i \in I} x_i n_{i2}, n_3(x) = \sum_{i \in I} x_i n_{i3}$$

and

$$b_1 = N_1, b_2 = N_2, b_3 = N_3.$$

A graphical representation of fuzzy constraint (2) is shown in Fig. 1.

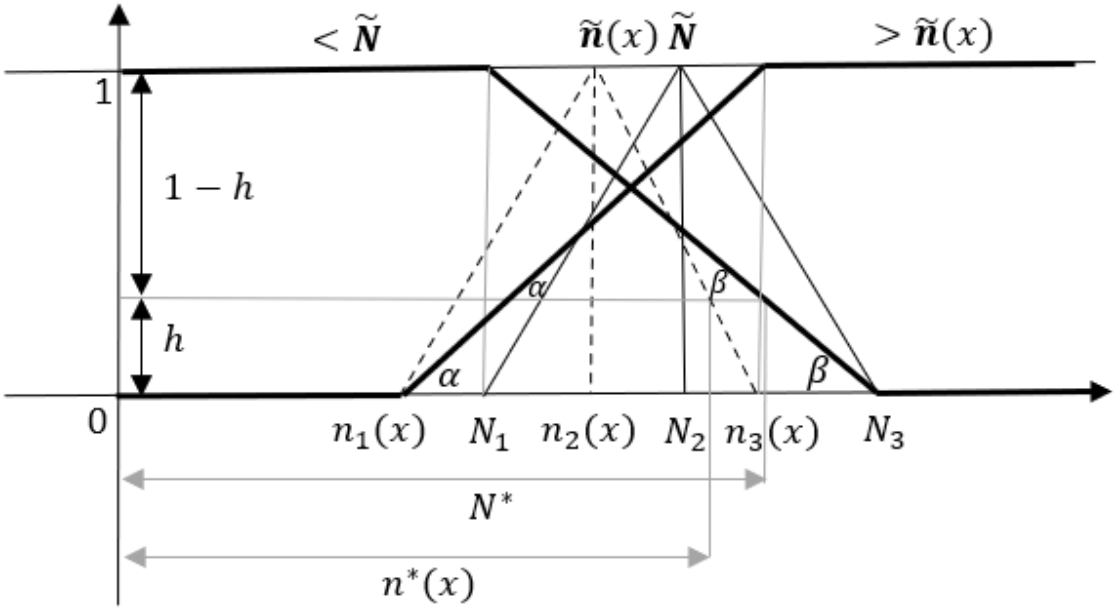


Figure 1 Graphical representation of fuzzy constraint parameters

Right Side Transformation

To the fuzzy number $\widetilde{\boldsymbol{N}}$ refer the triangles with vertices at points with coordinates $[N_1; 0]$, $[N_2; 1]$ a $[N_3; 0]$. Since the constraint is in the form $\widetilde{\boldsymbol{n}}(x) \leq \widetilde{\boldsymbol{N}}$, the progress highlighted in Figure $1 < \widetilde{\boldsymbol{N}}$, represents fuzzy values less than $\widetilde{\boldsymbol{N}}$. Using the similarity of triangles, it is also possible to express the relationship to the level of satisfaction based on coordinates $[N^*; h]$, $[N_1; h]$ and $[N_1; 1]$, which is a triangle similar to a triangle $[N_1; 0]$, $[N_1; 1]$ a $[N_3; 0]$. Value N^* can be then expressed from the similarity equation

$$tan\,\beta = \frac{1}{N_3 - N_1} = \frac{1-h}{N^* - N_1}$$

as follows:

$$N^* = N_1 + (1-h)(N_3 - N_1).$$

Left Side Transformation

In the case of the left side, it is a triangle $[n_1(x); 0]$, $[n_2(x); 1]$ and $[n_3(x); 0]$, which is similar to a triangle $[n_1(x); h]$, $[n_2(x); 1]$ a $[n_3(x); h]$. The fuzzy constraint of the left side can be conceived as a value that is greater than $\widetilde{\boldsymbol{n}}(x)$, inf the Figure $1 > \widetilde{\boldsymbol{n}}\,(x)$. This is to say that if we want to achieve the maximum level of satisfaction, the left side must be larger than the worst-case scenario in terms of the cost of implementing projects. Value $n^*(x)$ can be then expressed from the equation

$$tan\,\alpha = \frac{1-h}{n_3(x) - n^*(x)} = \frac{1}{n_3(x) - n_1(x)}$$

as follows:

$$n^*(x) = n_3(x) - (1-h) * (n_3(x) - n_1(x)).$$

After obtaining the values n^* and N^* the constraint (2), originally containing uncertainty, can be written in the form

$$n_3(x) - (1-h) * (n_3(x) - n_1(x)) \leq N_1 + (1-h)(N_3 - N_1)$$

Fuzzy Mathematical Model

The fuzzy mathematical model will take the form:

$$f(x,h) = h \rightarrow max \tag{5}$$

subject to:

$$n_3(x) - (1-h) * (n_3(x) - n_1(x)) \leq N_1 + (1-h)(N_3 - N_1) \tag{6}$$

$$\sum_{i \in I} \sum_{j \in J} a_{ij} x_i \geq h \cdot F^{max} + (1-h) \cdot (F^{max} - r) \tag{7}$$

(3)

$$h \in \langle 0; 1 \rangle \tag{9}$$

The objective function (5) ensures the requirement to maximize the level of satisfaction h. The approach principle based on the fuzzy mathematical model is not just maximization of indicators fulfilment, it maximizes the level of satisfaction to a sufficiently large value of the total fulfilling of indicators determined by the decision-maker. Constraint (6) is a transformation of constraint (2) into fuzzy form according to the logic given in the text in the Fuzzy parameters section. The constraint (7) specifies what numbers are large enough for the decision-maker at a given satisfaction level. Values F^{max} and r are determined by the requirements of the decision-maker and are based on predetermined requirements for fulfilment of the minimum values of the monitored indicators for the implementation period of the investment program. It is a value of sufficiently large values of cumulative filling of indicators and a tolerance expressing the degree of acceptability of lower values of cumulative filling of indicators. These two parameters can be used to evaluate the quality of the achieved solution. Link between satisfaction level h and the cumulative filling of the indicators is shown in Figure 2. The constraint (9) determines the definition scope of variables in the model.

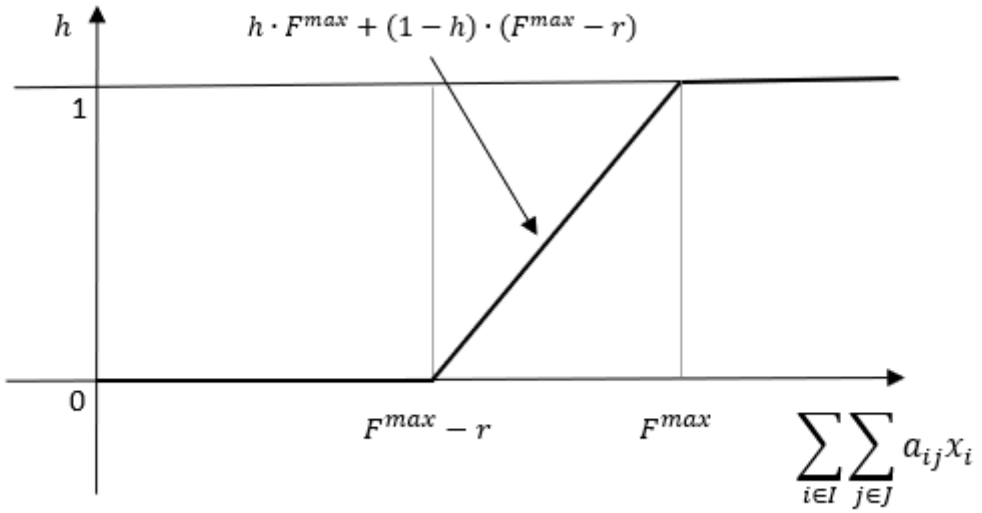


Figure 2 Relationship between the level of satisfaction h and cumulative filling of indicators

When solving a mathematical model (5) – (9), there is a problem with nonlinearity, which appears in the constraint (6), where occurs the multiplication of the variable h and parameters $n_1(x)$ and $n_3(x)$. The problem of nonlinearity of the model can be solved using the Tanaka-Asain iterative approach [7] based on the gradual increment of the parameter h value until the set of feasible solutions is emptied.

Mathematical model after variable transformation h will take the form (1), subject to (3), (6)-(7).

4 Computational Experiments

Computational experiments were performed on a sample of 10 model projects, which were generated based on real projects and adapted to the size of the workload. In addition, 3 monitored indicators were chosen. Values of F^{max} and r have been set as $F^{max} = 35$ and $r = 10$. Next, the initial value was set $h_0 = 0$ and $\Delta h = 0{,}01$.

Table 1 shows the input data and the outputs of the mathematical model.

Project ID	Cost			Indicator			
	n_{i1}	n_{i2}	n_{i3}	a_{i1}	a_{i2}	a_{i3}	x_i
1	15	15	18	16	7	11	0
2	4	4	6	5	2	2	0
3	5	5	6	5	2	3	0
4	6	6	8	8	3	4	1
5	1	1	3	1	3	3	1
6	6	7	10	6	1	4	0
7	3	3	4	4	1	2	0
8	5	5	7	7	3	3	1
9	5	5	8	4	2	3	0
10	5	5	6	8	1	1	0
In total optimized	12	12	18	16	9	10	

Table 1 Input and output data

Table 2 lists the results of the selected iterations, representing the different sets of selected projects, specifying the appropriate level of satisfaction representing the assumption that the available budget will not be exceeded.

Value of satisfaction	i where $x_i = 1$	Crisp objective function value $f(x)$
$h = 0$	$\{4; 5; 7; 8; 10\}$	52
$h = 0{,}25$	$\{4; 5; 8; 10\}$	45
$h = 0{,}5$	$\{2; 4; 5; 7\}$	38
$h = 0{,}72$	$\{4; 5; 8\}$	35

Table 2 Results of selected iterations

The table of outputs shows that the total filling of the indicators is 35, which directly corresponds to the value F^{max}. The optimization calculation was terminated at value $h = 0{,}72$.

5 Discussion of the Solution

Optimization calculations have reached the level of satisfaction $h = 0{,}72$. Therefore, the satisfaction level does not reach the maximum possible value of 1, even though the optimization calculation has reached the value F^{max}. This is caused due to the fact that compliance with the constraint cannot be guaranteed (6). In practice, this would mean that the limit of available funds could be exceeded. However, the selected combination of projects is still the best of all other combinations in terms of minimizing this threat.

It should be noted that the left- and right-side concept in constraint (6) is only one of the available options. This concept is determined by the decision-maker and is based on specific needs. For example, another approach might involve leaving the satisfaction level at its maximum level at n_{i2}, as these values belong to the fuzzy number at the highest level of satisfaction. The authors of the article tested this concept by a computational experiment, and the level of satisfaction reached $h = 1$, and 4 projects were selected instead of 3. However, this approach has not been found to be suitable for practical application in the current dynamic economic conditions.

6 Conclusion

The aim of the paper was to model the uncertainty that is burdened by the financial aspects of investments within the task of optimizing the selection of investment projects of transport infrastructure using fuzzy logic. Both the costs of implementing individual projects and the total amount of the available budget of the investment program are expressed as fuzzy. The values expressed this way describe real-world aspects more faithfully, modelling the uncertainty associated in particular with the development of the economic situation on the market.

According to the performed computational experiments, the chosen approach seems to be suitable for application to a data set representing a set of investment projects of transport infrastructure in order to select a portfolio of projects for implementation. The output can represent an important input into addressing the level of risk represented in the task of portfolio selection while maximizing the fulfilment of the values of the monitored indicators. From the point of view of the issue of maintaining the available budget while maximizing indicators, the decision-maker obtains important information about the level of credibility at which the various sets of projects representing a potential implementation portfolio will meet the presumption of not exceeding the budget even with uncertain monetary inputs.

The application of the approach seems to be particularly suitable in the case of selecting a portfolio of larger projects whose preparation takes a longer period of time, because, as it often turns out in practice, the several-year preparation time of large investment projects in the currently unstable economic situation and dynamic changes on the market makes it difficult to accurately determine the monetary parameters of the upcoming projects and investment programs in the implementation time horizon in the initial phase of implementation. The approach can be applied when the decision-maker is able to determine at the expert level possible scenarios for the development of the values of investment requirements of projects and the volume of available funds in the investment programme.Acknowledgements

Acknowledgements

This work was supported by project SGS22/126/OHK2/2T/16 Design of Computational Methods for Optimizing Investment Projects Portfolio in Transport Infrastructure.

References

[1] Bellman, R. E., & Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management science* 17(4), 141-164. https://doi.org/10.1287/mnsc.17.4.B141

[2] Çelikbilek, Y. & Tuysuz, F. (2015). A Fuzzy Multi Criteria Decision Making Approach for Evaluating Renewable Energy Sources. *The 4th International Fuzzy Systems Symposium Proceedings.*

[3] European Comission, Directorate-General for Economic and Financial Affairs, Kalantzis, F., Arnoldus, P., Maincent, E., et al. (2015). *Infrastructure in the EU: developments and impact on growth*. Publications Office. https//:doi.org/doi:10.2765/85301.

[4] Hong, J., Chu, Z. & Wang, Q. (2011). Transport infrastructure and regional economic growth: Evidence from China. *Transportation* 38, 737-752. https://doi.org/10.1007/s11116-011-9349-6.

[5] Klir, G. J. & Yuan, B. (2015). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall.

[6] Sanchez-Robles, B. (1998). Infrastructure Investment and Growth: Some Empirical Evidence. *Contemporary Economic Policy* 16(1), pp. 98-108.

[7] Tanaka, H. & Asai, K. (1984). Fuzzy linear programming problems with fuzzy numbers, *Fuzzy Sets and Systems* 13, 1-10. ISSN 0165-0114. https://doi.org/10.1016/0165-0114(84)90022-8.

[8] Zadeh, L. A. (1975) Fuzzy logic and approximate reasoning. *Synthese* 30, 407–428. https://doi.org/10.1007/BF00485052

[9] Zaeimi, M. B. & Rassafi, A. A. (2021). Designing an integrated municipal solid waste management system using a fuzzy chance-constrained programming model considering economic and environmental aspects

under uncertainty, *Waste Management* 125, 268-279. ISSN 0956-053X. https://doi.org/10.1016/j.wasman.2021.02.047.

Threshold Values for Calculating the Efficiency of a Transport Company's Investment

Petr Jiříček[1], Stanislava Dvořáková[2]

Abstract. The aim of the contribution is to determine the threshold values of the parameters of the evaluation of the efficiency of the investment project of a transport company using the methods of net present value and internal rate of return. The simulation of the railway transport project will be set for seasonal and year-round operation of the railway in the alternatives of a non-inflationary discount rate, a discount rate corresponding to an increased rate of inflation and subsequently under the conditions of the current high rate of inflation. The effect of any subsidy from public funds on the threshold values of the project's effectiveness will also be considered.

Keywords: Threshold value, investment projects, transport company, rate of inflation

JEL Classification: C20, H43
AMS Classification: 65H04

1 Introduction

In economic theory and practice, the threshold value of a investment project is conceived as a significant part of investment project efficiency evaluation. In this context, a beneficial impact of an investment in the operational stage means that sufficient means for investment costs payment are generated and an overall positive effect occurs after reaching the break-even point US EPA [10]. The classical break-even point analysis in the accounting conception provides information on the value of a variable determined in an accrual way (income, expenses). It means that we determine such amount of production or services at which the product or service sales revenues cover production or services costs. If we evaluate the investment efficiency based on cash-flow principle, we accept the definition of the so-called cash break-even point. In this concept, the break-even point is the moment when cash income from the investment (inflows) equals the expenditures related to the investment (outflows).

The modern approach to investment efficiency evaluation uses discounted cash-flows, i.e. it considers a different time value of money in individual stages of the investment process [1]. The most often used method for evaluating the economic efficiency of investments is the Net Present Value method [9]. The other method is the Internal Rate of Return [5]. Internal Rate of Return is not strongly NPV-consistent and financial analysis of project may even turn out to be impossible [8].

In this concept, cash-even point can be defined as the so called financial break-even point, i.e. the moment when the Net Present Value (NPV) of the evaluated investment equals zero [2]. Another approach to determining break-even point while meeting the requirement NPV=0 is offered by the definition of threshold value cash-flow. It is determining such marginal value of cash-flow in individual years of the investment cycle that causes an indifferent value of utility from the project implementation [6]. We will use this procedure to determine the threshold values of the investment project aimed at the development of a regional railway line in the territory of the Vysočina Region. For the calculation, we will use available data from the cost-benefit analysis study of this project, prepared for the Vysočina Region.

2 Objective and Methods

The presented paper aims at calculating break-even point by means of modelling the threshold value cash-flow in individual years of the investment cycle of a real transport project. It is based on a case study of an investment project of regional railway line [7]. To define the model, we will use the European Union methodology for evaluation of investment projects financed from the European funds for the 2014–2020 cohesion period, European Commission [4]. The track operator reports data for the so-called economic year, which is also presented in the above-mentioned study of the cost-benefit analysis of the project. The model thus includes already established

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, petr.jiricek@vspj.cz
[2] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz

investment and operating costs for the period from 6/2020 to 6/2022, and subsequently predicted investment and operating costs for the period until 6/2030.

Because the railway line development project considers two basic scenarios: year-round operation or seasonal operation, the threshold value simulation will follow up on the defined project scenarios. For the presented paper, two further defined scenarios will be modelled:

I. Year-round operation of railway line

II. Seasonal operation of railway line

Every from these scenarios will contain 3 alternatives depending on the level of the inflation rate:

A. Low inflation rate alternative (corresponds to the simulation discount rate of the investment under the conditions set in the European Directive, i.e. 4 % p.a.)

B. Higher inflation rate alternative (corresponds to the simulation of a 7% discount rate of the investment)

C. The alternative of current high rate of inflation (corresponds to the simulation of a 15% discount rate of the investment)

The Net Present Value method (NPV) states what cash flow caused by the investment is left after deducting investment costs in pre-projected lifetime

$$NPV = \sum_t \frac{CF_t}{(1+k)^t} \tag{1}$$

where CF_t is the cash flow of the transport project, i.e. the inflows minus outflows of the project,

k is the discount rate (the required return on investment from the perspective of the investor of the project),

NPV is net present value of the transport project reflecting the benefit from the project from a view of the given investment.

This relation can be defined as the project investment curve [3]. The Internal Rate of Return of the project (IRR) is defined as the root of the investment curve (1) of the project, i.e.

$$NPV = \sum_t \frac{CF_t}{(1+IRR)^t} = 0. \tag{2}$$

3 Results and Discussions

Statuses and predictions of the development of operating and investment cash-flows of the regional rail-way line for the considered period of the project (the economic year always starts in June of the previous year) are in Table 1 for year-round operation of railway line and in Table 2 for seasonal operation of railway line.

date	Investment CF	Operating CF
6/2021	–365	1 013
6/2022	–120	–3 637
6/2023	–1 862	–109
6/2024	–1 243	1 013
6/2025	–1 411	–3 637
6/2026	–447	–109
6/2027	–39	1 013
6/2028	–20	–3 637
6/2029	–102	–109
6/2030	–365	1 013

Source: Vysočina Region, 2022

Table 1 State and prediction of the development of annual operating and investment CF (year-round operation) (in thousands of CZK)

date	Investment CF	Operating CF
6/2021	–365	202.6
6/2022	–120	–727.4
6/2023	–1 862	–21.8
6/2024	–1 243	202.6
6/2025	–1 411	–727.4
6/2026	–447	–21.8
6/2027	–39	202.6
6/2028	–20	–727.4
6/2029	–102	–21.8
6/2030	–365	202.6

Source: Vysočina Region, 2022

Table 2 State and prediction of the development of annual operating and investment CF (seasonal operation) (in thousands of CZK)

3.1 Calculation of the Efficiency Values of the Railway Line Project for Individual Operating Scenarios Using the NPV Method

Now we will calculate the Net Present Value values for individual alternatives of the discount rate (corresponding to the amount of the inflation rate) for the case of both defined railway operation scenarios. To calculate the function $NPV = f(k)$ for the considered years of the project, we apply formula (1):

$$NPV = \sum_{t=-2}^{7} \frac{CF}{(1+k)^t} = CF_{-2}(1+k)^2 + CF_{-1}(1+k) + CF_0 + \frac{CF_1}{1+k} + \frac{CF_2}{(1+k)^2} + \cdots \frac{CF_7}{(1+k)^7} \tag{3}$$

Criteria:

$NPV > 0$ – The investment yields positive utility

$NPV < 0$ – The investment yields negative utility

$NPV = 0$ – The investment yields zero utility

	Scenario I. *Year-round operation of railway*	**Scenario II.** *Seasonal operation of railway*
k=4%	–12 407	–7 031
k=7%	–11 928	–6 791
k=15%	–10 926	–6 295

Table 3 NPV values for various alternatives of railway operation (in thousands of CZK)

Commentary

The results of the calculation of the efficiency of the railway line investment using the NPV method from 6/2021 (years –2, –1 of the pre-investment phase and years 0, 1, 2, 3, 4, 5, 6, 7 of the project prediction in model) show us that the project is unprofitable. It will be necessary to simulate cash-flow in individual years, so that it is possible to reach a cash break-even point at which the project shows a simple return (i.e. $NPV = 0$.)

In the following Figures 1, 2 the NPV curves for the above-mentioned scenarios I. and II. are displayed. In detail, we then see the NPV values for the individual chosen discount rates of the project.

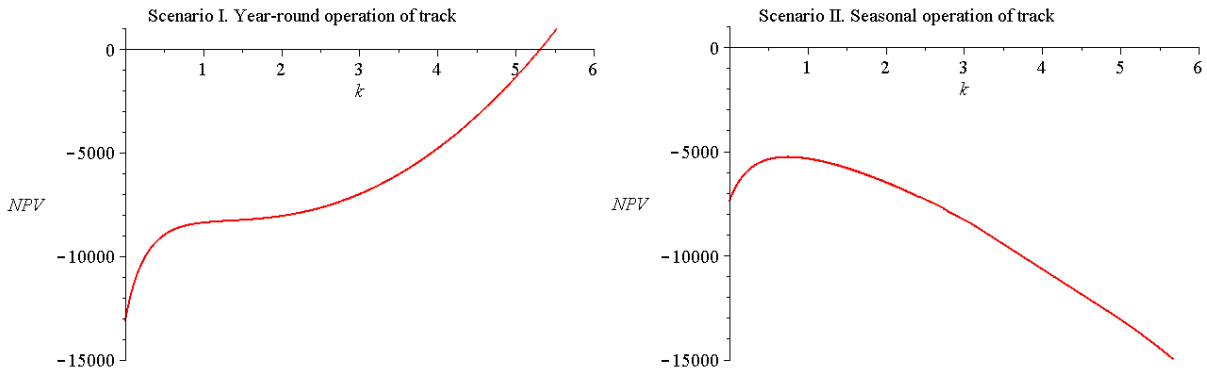


Figure 1 Investment curves $NPV = f(k)$ for individual scenarios

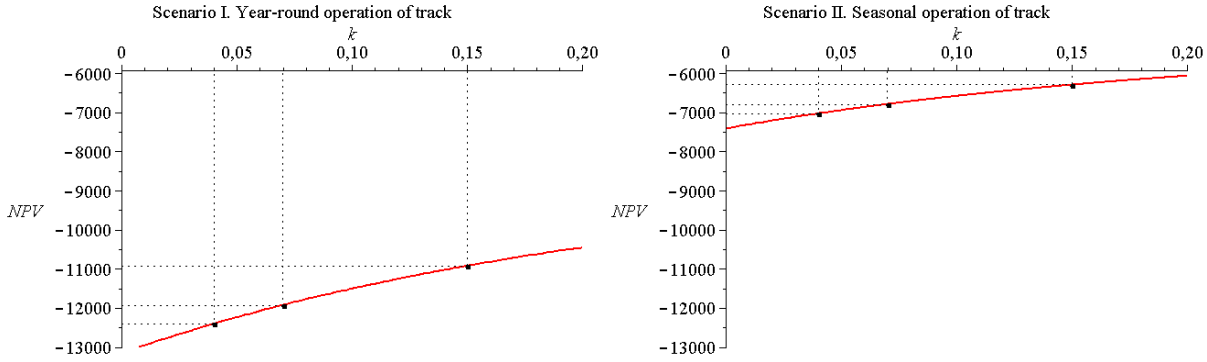


Figure 2 Investment curves $NPV = f(k)$ for individual scenarios (details)

3.2 Calculation of the Efficiency Values of the Railway Line Project for Individual Operating Scenarios Using the IRR Method

To calculate the root of the function NPV = f(k) for the considered years of the project, we apply the formula (2):

$$NPV = CF_{-2}(1+IRR)^2 + CF_{-1}(1+IRR) + CF_0 + \frac{CF_1}{1+IRR} + \frac{CF_2}{(1+IRR)^2} + \cdots \frac{CF_7}{(1+IRR)^7} = 0 \qquad (4)$$

Scenario I.	**Scenario II.**
Year-round operation of railway	*Seasonal operation of railway*
532%	∞

Table 4 IRR values for different railway operation scenarios

Commentary

From the results of the IRR method in Table 4 and from Figures 1 and 2 we can see that the IRR method cannot be used to evaluate the effectiveness of the project. The IRR method gives us a result correspond-ing to the economic evaluation of the investment only if the curve $NPV = f\ (k)$ is monotonically decreasing in the field of positive real numbers.

3.3 Determination of Threshold Values of the Railway Line Project for Individual Operation Scenarios

Now we will examine the threshold values of input variables, i.e. cash flows of the project for the selected discount rates. From formula (3) we can draw the general formula for calculation of threshold value in individual years of the investment cycle, i.e. the variable CF_i (for $i = -2, \ldots, 7$) while in our case we consider the condition NPV=0 and the discount rate $k = 0.05$; $k = 0.07$ and $k = 0.15$:

$$CF_{-2}(1+k)^2 + CF_{-1}(1+k) + (CF_0 + Y) + \frac{CF_1 + Y}{1+k} + \frac{CF_2 + Y}{(1+k)^2} + \cdots + \frac{CF_7 + Y}{(1+k)^7} = 0, \qquad (5)$$

where Y required constant increase in CF in all years of operation so that the condition holds $NPV = 0$.

The model for calculation of threshold value will now be tested on a real proposed railway line moderni-zation project financed from the European funds, which will be proposed to be implemented in years **2023–2030**. In following Table 5 we will find out required constant increase in CF in all years of operation.

Discount rate	**Scenario I.** *Year-round operation of railway*	**Scenario II.** *Seasonal operation of railway*
k=4%	1 771.9	1 004.1
k=7%	1 866.9	1 062.9
k=15%	2 117.3	1 220.0

Table 5 Resulting values of Y for the given railway operation scenarios for different rates of inflation (in thousands of CZK)

In Table 6 and Table 7 we will find out the threshold values of critical variables of projects, i.e. such values of cash flows in all years of model project scenarios when the net present value of the project equals zero and the project does not provide utility. The *threshold value* thus shows the critical value of cash flows in individual years of the project lifetime for given scenarios.

date	**k=4%**	**k=7%**	**k=15%**
6/2021	–162.4	–162.4	–162.4
6/2022	–847.4	–847.4	–847.4
6/2023	–879.7	183.2	1403.2
6/2024	–36.3	1026.6	2246.6
6/2025	–1134.3	–71.4	1148.6
6/2026	535.3	1598.2	2818.2
6/2027	1167.7	2230.6	3450.6
6/2028	256.7	1319.6	2539.6
6/2029	880.3	1943.2	3163.2
6/2030	841.7	1904.6	3124.6

Table 6 Prediction of threshold value in case of year-round operation of railway (in thousands of CZK)

date	**k=4%**	**k=7%**	**k=15%**
6/2021	648	648	648
6/2022	–3757	–3757	–3757
6/2023	–199.1	1667.8	3785.1
6/2024	1541.9	3408.8	5526.1
6/2025	–3276.1	–1409.2	708.1
6/2026	1215.9	3082.8	5200.1
6/2027	2745.9	4612.8	6730.1
6/2028	–1885.1	–18.2	2099.1
6/2029	1560.9	3427.8	5545.1
6/2030	2419.9	4286.8	6404.1

Table 7 Prediction of threshold value in case of seasonal operation of railway (in thousands of CZK)

Based on the results in Tables 5–7 we can state that the threshold value in examined model scenarios (year-round operation of railway, seasonal operation of railway). The cash-flow values for individual discount rate alternatives document the adverse effect of higher inflation rates on the course of the project.

4 Conclusion

Determining the threshold value of a public investment project can be defined as such a value of cash flow in individual years of the project cycle when the cash break-even point is reached. It is the situation when zero utility

of the project is reached as measured by net present value. The threshold value determines the sensitivity of the output variable of the project to the change of the input variable. Based on the results in Tables 6, 7 we can say that the threshold value in simulated model scenarios (year-round operation of rail-way, seasonal operation of railway). documents the adverse effect of higher inflation rates on the course of the project.

References

[1] Boardman, A., Greenberg, D., Vining, A. & Weimer, D. (2014). Cost-Benefit Analysis. *Concepts and Practise*. 4th edition. Harlow: Pearson Education Limited.

[2] Brealey, R., Myers, S. & Allen, F. (2011). *Principles of Corporate Finance*. 2th. edition. New York: McGraw-Hill Irwin.

[3] Dvořáková, S. & Jiříček, P. (2013): Modelling Financial Flows of Development Projects Subsidized from European Funds. *Proceedings of the 31st International Conference on Mathematical Methods in Economics*. (pp. 119–128). Jihlava: Vysoká škola polytechnická Jihlava.

[4] European Commission (2015). *Guide to Cost-Benefit Analysis of Investment Project*. Brusel: DG Regional and Urban Policy.

[5] Hazen G. (2003). A new perspective on multiple internal rates on return. *The Engineering Economist* 48(1), 31–51.

[6] Jiříček, P. & Dvořáková, S. (2017). Investment Projects Threshold Value Simulation. *Proceedings of the 35rd Mathematical Methods in Economics* (pp. 301–306). Hradec Králové: University of Hradec Králové.

[7] Kraj Vysočina (2022). *Cost-benefit analýza projektu "Úzkokolejka Jindřichův Hradec-Obrataň"*. Jihlava: Vysoká škola polytechnická Jihlava.

[8] Magni, C. & Marchioni, A. (2020). Average rates of return, working capital, and NPV-consistency in project appraisal: A sensitivity analysis approach. *International Journal of Production Economic 229 (11)*, 107769, https://doi.org/10.1016/j.ijpe.2020.107769

[9] Osborne, M. (2010). A resolution to the NPV-IRR debate? *The Quarterly Review of Economics and Finance 50 (2)*, 234–239

[10] United States Environmental Protection Agency. (2009). *Guidance on the Development, Evaluation, and Application of Environmental Models 2009*. [online]. Washington: EPA/100/K-09/003. Available at: www.epa.gov/crem [cited 2017-12-03]

Estimation of the General Measure of Stochastic Non-Dominance

Jana Junová[1]

Abstract. This paper deals with the general measure of stochastic non-dominance, which was developed to understand the situations when stochastic dominance rules are broken. It can be derived analytically for some distributions. However, in the case of other distributions, such as the log-normal and gamma, it is useful to be able to estimate it. We do so by a numerical study, in which we approximate these distributions by discrete distributions with equiprobable atoms. This paper describes how the numerical study is designed, and presents the estimated measures of stochastic non-dominance between selected distributions.

Keywords: stochastic dominance, measure of stochastic non-dominance, approximation

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Stochastic dominance is a concept that enables the comparison of investment opportunities without the precise knowledge of a particular investor's preferences. It accepts the fact that each investor may have a different utility function and works with whole classes of them. It allows us to compare investments under the only assumption that the correct utility function is in a particular class of them. The widest class includes all non-decreasing and continuous utility functions. It corresponds to the first-order stochastic dominance (FSD), which was introduced by [15]. Second-order stochastic dominance (SSD) introduced by [2], [16], and [17] allows the comparison of investments under the additional assumption that the investor's utility function is concave.

We follow [8] in defining stochastic dominance of the first and the second order.

Definition 1 (Stochastic Dominance)**.** We define the following sets of utility functions:
$U_1 = \{u \text{ utility function}, u' \geq 0\}$
$U_2 = \{u \text{ utility function}, u' \geq 0, u'' \leq 0\}$.
For $n \in \{1, 2\}$, we say that a random variable X dominates a random variable Y by the n^{th}-order stochastic dominance ($X \succeq_{(n)} Y$) if

$$\mathbb{E}\, u(X) \geq \mathbb{E}\, u(Y) \text{ for all } u \in U_n \text{ such that these expected values exist.}$$

Properties of stochastic dominance have been widely studied. Results regarding stochastic dominance in particular distributions can be found for example in [1] and [8]. An important application of stochastic dominance lies in portfolio optimization. This has been recently studied for example in [10], [4], [9], and [18].

The central focus of our work is to analyze the situations when the rules for stochastic dominance in particular distributions are broken. To bring more understanding to these situations, [3] have defined the general and the specific measures of stochastic non-dominance, which quantify how much the stochastic dominance rules are broken. They can be found by solving an optimization problem, which minimizes the Wasserstein distance between certain variables. The properties of the Wasserstein distance are presented in [12].

[3] have shown how the specific and in some cases also the general measures of non-dominance can be computed analytically between random variables with uniform, normal, or exponential distribution. However, their approach cannot be applied to some widely used distributions with more complicated distribution functions. This paper presents a way to estimate the general measure of non-dominance in these cases by a numerical study. We approximate the distributions of interest by discrete distributions with equiprobable atoms and compute the measure of stochastic non-dominance between them.

[1] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, Prague, 186 75, Czech Republic, junova@karlin.mff.cuni.cz

The definitions and some properties of the measures of stochastic non-dominance and the estimating procedure is presented in Section 2. In Section 3, the results of the numerical study as applied to log-normal and gamma distributions are presented.

2 Estimation Procedure

The general measure of stochastic non-dominance was defined by [3] in the following way.

Definition 2 (General Measure of Stochastic Non-Dominance)**.** Let X and Y be integrable random variables, and suppose that $\text{dist}(X, \hat{X})$ is the Wasserstein distance between the considered random variables. The general measure of n$^{\text{th}}$-order stochastic non-dominance between X and Y, $\text{GND}_n(X, Y)$, is computed by solving the following program:

$$\begin{aligned} \text{GND}_n(X, Y) = \min_{\hat{X} \in L_1} & \text{dist}(X, \hat{X}) \\ \text{subject to } & \hat{X} \succeq_{(n)} Y. \end{aligned} \tag{1}$$

It can seen that if $X \succeq_{(n)} Y$, then $\hat{X} = X$ and $\text{GND}_n(X, Y) = 0$. [3] have defined also the specific measure of stochastic non-dominance, which will be utilized in the numerical study.

Definition 3 (Specific Measure of Stochastic Non-Dominance)**.** Let X and Y be integrable random variables from a particular distribution family $\mathcal{M}$, and suppose that $\text{dist}(X, \hat{X})$ is the Wasserstein distance between the considered random variables. The specific measure of n$^{\text{th}}$-order stochastic non-dominance between X and Y, $\text{SND}_n(X, Y)$, is computed by solving the following program:

$$\begin{aligned} \text{SND}_n(X, Y) = \min_{\hat{X} \in \mathcal{M}} & \text{dist}(X, \hat{X}) \\ \text{subject to } & \hat{X} \succeq_{(n)} Y. \end{aligned} \tag{2}$$

[3] use the Wasserstein distance of order 2 to measure the distance between random variables X and $\hat{X}$. One needs to be able to compute the squared integrated difference of quantile functions of X and $\hat{X}$. This allows the precise analytical derivation of the values of SND_n and GND_n for some distributions. However, the quantile functions can be very complicated or may not even have a closed-form expression, which prevents us from deriving the measures of non-dominance analytically for some distributions.

In these cases, it becomes useful to be able to estimate them. We will estimate the distribution of interest by discrete distributions with equiprobable atoms (empirical distributions). The Wasserstein distance, and as a result also the specific measure of non-dominance can be computed very well between empirically distributed random variables. We will use the following theorem presented in [3] in the estimation procedure.

Theorem 1. *Let X and Y be discrete random variables with ordered equiprobable atoms $x_1 \leq \cdots \leq x_T$, $y_1 \leq \cdots \leq y_T$. We assume that $\hat{X}$ is also empirically distributed with atoms $\hat{x}_1, \ldots, \hat{x}_T$. Then the specific measure of n^{th}-order non-dominance between X and Y is computed, using the Wasserstein distance of integer order $r \geq 1$, as follows:*

$$\begin{aligned} \text{SND}_n(X, Y)^r = \min_{\hat{x}_1, \ldots, \hat{x}_T \in \mathbb{R}} & \frac{1}{T} \sum_{t=1}^{T} \mid x_t - \hat{x}_t \mid^r \\ \textit{subject to } & \hat{X} \succeq_{(n)} Y. \end{aligned} \tag{3}$$

The condition $\hat{X} \succeq_{(n)} Y$ can be formulated very easily for $n = 1$ and $n = 2$ as $\hat{x}_t \geq y_t$ for all t, or $\sum_{j=1}^{t} \hat{x}_j \geq \sum_{j=1}^{t} y_j$ for all t respectively. The condition $\hat{X} \succeq_{(n)} Y$ becomes significantly more complicated for higher orders of n. We will therefore focus on the measure of the first-order and the second-order non-dominance.

$\text{GND}_n(X, Y)$ will be estimated by the following procedure. We randomly generate T numbers from the distributions of X and T numbers from the distribution of Y, and use them as atoms of two empirical distributions. We denote these empirically distributed random variables, which approximate X and Y, as $\tilde{X}$ and $\tilde{Y}$. Using Theorem 1, we compute the $\text{SND}_n(\tilde{X}, \tilde{Y})$. We repeat this procedure k times. We compute the average $\text{SND}_n(\tilde{X}, \tilde{Y})$ and approximate the $\text{GND}_n(X, Y)$ by it.

Note that the specific measure of non-dominance between the empirically distributed $\tilde{X}$ and $\tilde{Y}$ approximates the general, not the specific measure of non-dominance between the original variables. The fact that the hypothetical variable, which dominates $\tilde{Y}$, is empirically distributed does not ensure that it approximates a random variable from the original distribution family of X and Y.

3 Results of the Numerical Study

We present the results of the estimations of the general measure of stochastic non-dominance between random variables with log-normal and gamma distribution. The quantile function of log-normal distribution is very complicated. To our best knowledge and according to [11], there is no closed-form expression for the quantile function of gamma distribution. We, therefore, are not able to derive an exact formula for the computation of the measures of non-dominance between random variables with these distributions and proceed by estimating them. We approximate the distributions of interest by empirical distributions with $T = 1000$ atoms. For each pair of the considered distributions, the approximation procedure is repeated $k = 200$ times.

3.1 Log-Normal Distribution

Log-normal distribution is a distribution, which is in a tight relationship with normal distribution. If a random variable X is log-normally distributed, then random variable $Z = \ln(X)$ is normally distributed. The distribution is determined by two parameters, μ and σ. μ is a real number, σ is a positive real number. Its probability density function is

$$\frac{1}{x\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \cdot \mathbb{I}_{x>0}.$$

Its mean is $\eta = e^{\mu+\sigma^2/2}$ and its variance is $s^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

It was shown in [8] that the following rules hold for stochastic dominance in log-normal distributions hold.

Proposition 2. *Suppose η_X, η_Y, s_X, s_Y are the means and standard deviations of X and Y. Then*

$$X \succeq_{(1)} Y \iff \eta_X \geq \eta_Y \text{ and } s_X^2 = s_Y^2,$$

$$X \succeq_{(2)} Y \iff \eta_X \geq \eta_Y \text{ and } \frac{\eta_X}{s_X} \geq \frac{\eta_Y}{s_Y}.$$

We suppose for the purpose of the numerical study that $X \sim ln(0, 1)$ and we alternate the parameters of Y. For each set of parameters of Y, we estimate the $\mathrm{GND}_n(X, Y)$ by the described procedure. We present the mean estimated $\mathrm{GND}_n(X, Y)$ as well as the standard deviation of such estimate in Table 1. The first two columns specify the parameters defining Y, the third column describes the strongest holding stochastic dominance relationship if $X \succeq Y$, and the following columns present the average $\mathrm{SND}_1(\tilde{X}, \tilde{Y})$ and $\mathrm{SND}_2(\tilde{X}, \tilde{Y})$ and their standard deviations for each set of parameters.

It can be seen from the results presented in Table 1 that GND_2 is always lower than (or equal to) GND_1. It should be so because FSD is stricter than SSD so $\hat{X}$ has to satisfy stricter conditions in order to dominate Y with respect to FSD. As a result, it differs from X more and the measure of non-dominance is higher.

When $X \succeq_{(2)} Y$, the estimated GND_2 is 0 as we expect. Both GND_1 and GND_2 increase with increasing μ_Y. It is understandable because $\eta_X \geq \eta_Y$ is a necessary condition for both $X \succeq_{(1)} Y$ and $X \succeq_{(2)} Y$, and $\eta_X \geq \eta_Y \iff e^{\mu_X+\sigma_X^2/2} \geq e^{\mu_Y+\sigma_Y^2/2}$. If $\sigma_X = \sigma_Y$, it implies that $\mu_X \geq \mu_Y$ is a necessary condition for $X \succeq_{(1)} Y$ and $X \succeq_{(2)} Y$. As μ_Y increases, μ_X becomes increasingly far from being higher than μ_Y.

A similar rule holds for σ^2. With increasing σ_Y^2, both GND_1 and GND_2 increase. The only exception to this rule can be seen by comparing the first two rows of Table 1, where the GND_2 decreases to zero. It is correct in this case because SSD does not hold for the parameters defining the first row.

3.2 Gamma Distribution

Gamma distribution can be seen as a generalization of exponential distribution. It uses two parameters, $k > 0$ and $\theta > 0$, and its probability density function is

$$\frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \cdot \mathbb{I}_{x>0}.$$

μ_Y	${\sigma^2}_Y$	SD holding	GND_1	std GND_1	GND_2	std GND_2
-1	0.25	-	0.003	0.001	0.003	0.001
-1	1	SSD	0.003	0.035	0.0	0.0
-1	2	SSD	0.814	0.931	0.0	0.0
-1	4	-	11.281	7.14	1.01	0.476
0	0.25	-	0.136	0.014	0.136	0.014
0	2	-	4.547	2.277	1.025	0.219
0	4	-	36.392	28.919	5.622	1.467
1	0.25	-	1.541	0.064	1.516	0.06
1	1	-	4.822	0.789	2.838	0.188
1	2	-	16.324	4.611	5.718	0.6
1	4	-	101.44	63.264	18.455	3.868
2	0.25	-	7.252	0.202	6.749	0.156
2	1	-	17.223	1.907	10.514	0.536
2	2	-	48.474	12.898	18.442	1.641
2	4	-	302.706	236.554	53.8	11.572

Table 1 Estimated GND in log-normal distributions with reference distribution $ln(0, 1)$.

Gamma distribution with parameters $k = 1$ and $\theta = \frac{1}{\lambda}$ is equivalent to exponential distribution with parameter λ. It holds for $X \sim \Gamma(k, \theta)$ that $\mathbb{E}\, X = k\theta$ and $\text{var}\, X = k\theta^2$.

Proposition 3 ([1]). *Let $X \sim \Gamma(k_X, \theta_X)$ and $Y \sim \Gamma(k_Y, \theta_Y)$. Then*

$$X \succeq_{(1)} Y \iff \theta_X \geq \theta_Y \text{ and } k_X \geq k_Y,$$

$$X \succeq_{(2)} Y \iff \frac{k_X}{k_Y} \geq \frac{\theta_Y}{\theta_X} \text{ and } k_X \geq k_Y.$$

The study proceeds in the way that was described above. First, we set $X \sim \Gamma(1, 1)$ and alternate the parameters defining Y. We present the results for $\theta_Y = 1, 2, 3$ and $k_Y = 1, 2, 3, 4$ in Table 2. The first two columns specify the parameters defining Y, and the following columns present the average $\text{SND}_1(\tilde{X}, \tilde{Y})$ and $\text{SND}_2(\tilde{X}, \tilde{Y})$ and their standard deviations for each set of parameters. In case of the presented parameters $X \not\succeq_{(n)} Y$ for any n.

k_Y	θ_Y	GND_1	std GND_1	GND_2	std GND_2
1	2	1.431	0.096	1.006	0.064
1	3	2.805	0.152	1.986	0.096
2	1	1.093	0.063	0.997	0.05
2	2	3.524	0.129	3.002	0.1
2	3	5.977	0.181	5.008	0.136
3	1	2.148	0.071	2.002	0.061
3	2	5.59	0.141	5.002	0.118
3	3	9.04	0.2	7.993	0.163
4	1	3.183	0.088	3.004	0.079
4	2	7.634	0.152	7.006	0.133
4	3	12.104	0.217	11.002	0.183

Table 2 Estimated GNDs in gamma distributions with reference distribution $\Gamma(1, 1)$.

It can be seen in Table 2 that $\text{GND}_1(X, Y)$ is higher than $\text{GND}_2(X, Y)$ for all presented parameters, which is correct. Both $\text{GND}_1(X, Y)$ and $\text{GND}_2(X, Y)$ increase with increasing k_Y or θ_Y. It makes sense because as k_Y increases, $k_X = 1$ is even further from being higher than k_Y, which is a necessary condition for $X \succeq_{(n)} Y$. Similarly, $\theta_X \geq \theta_Y$ is a necessary condition for $X \succeq_{(1)} Y$, and increasing θ_Y leads to $\theta_X = 1$ being further from satisfying it. It holds for fixed k_X and θ_X that if θ_Y increases, it becomes even further from satisfying also $X \succeq_{(2)} Y$. We may also notice

that if the dominance is violated only by the value of k_Y ($\theta_Y = 1$), the measures of non-dominance tend to be lower than when the dominance is violated only by the value of θ_Y ($k_Y = 1$).

The distributions defined by the first two rows of Table 2 where $k_Y = 1$ can be seen as exponential distributions with parameters $\lambda_Y = 1/2$ and $\lambda_Y = 1/3$. The reference distribution $\Gamma(1, 1)$ is also an exponential distribution with parameter $\lambda_X = 1$. According to the following theorem proved by [3], it is possible to compute the general measure of the first-order non-dominance in these two cases exactly.

Theorem 4. *The general measure of stochastic non-dominance between two exponentially distributed random variables, $X \sim Exp(\lambda_X)$, $Y \sim Exp(\lambda_Y)$, is the following:*

1. *If $\lambda_X \leq \lambda_Y$, then* $\mathrm{GND}_1(X, Y) = 0$.
2. *If $\lambda_X > \lambda_Y$, then* $\mathrm{GND}_1(X, Y) = \sqrt{2} \cdot \frac{\lambda_X - \lambda_Y}{\lambda_X \lambda_Y}$.

For $\lambda_X = 1$ and $\lambda_Y = \frac{1}{2}$, the $\mathrm{GND}_1(X, Y)^2$ is $2 \cdot (1 - \frac{1}{2})^2/(1 \cdot \frac{1}{2})^2 = 2$. For $\lambda_X = 1$ and $\lambda_Y = \frac{1}{3}$, the $\mathrm{GND}_1(X, Y)^2$ is $2 \cdot (1 - \frac{1}{3})/(1 \cdot \frac{1}{3})^2 = 8$. Their square roots are approximately 1.414 and 2.828. These values correspond quite well to the estimated $\mathrm{GND}_1(X, Y)$.

We show the estimated measures of non-dominance for another set of gamma distributions. This time $X \sim \Gamma(2, 1)$. We alternate the parameters defining Y: $\theta_Y = 1, 2, 3$ and $k_Y = 1, 2, 3, 4$. The numerical study follows the same pattern. The results are shown in Table 3. The first two columns specify the parameters defining Y, the third column describes the strongest holding stochastic dominance relationship if $X \succeq Y$, and the following columns present the average $\mathrm{SND}_1(\tilde{X}, \tilde{Y})$ and $\mathrm{SND}_2(\tilde{X}, \tilde{Y})$ and their standard deviations for each set of parameters.

k_Y	θ_Y	SD holding	GND_1	std GND_1	GND_2	std GND_2
1	1	FSD	0.002	0.008	0.0	0.0
1	2	SSD	0.575	0.119	0.033	0.049
1	3	-	1.892	0.176	1.0	0.11
2	2	-	2.454	0.134	2.005	0.099
2	3	-	4.906	0.195	4.0	0.149
3	1	-	1.062	0.074	1.001	0.063
3	2	-	4.496	0.141	3.998	0.111
3	3	-	7.929	0.188	6.974	0.156
4	1	-	2.102	0.086	2.007	0.076
4	2	-	6.551	0.164	6.014	0.141
4	3	-	10.986	0.206	9.987	0.181

Table 3 Estimated GNDs in gamma distributions with reference distribution $\Gamma(2, 1)$.

Firstly, we may notice in Table 3 that when $X \succeq_{(1)} Y$, both $\mathrm{GND}_1(X, Y)$ and $\mathrm{GND}_2(X, Y)$ either equal 0, or they are very close to it, which can be attributed to the simulation error. When $X \succeq_{(2)} Y$, $\mathrm{GND}_2(\tilde{X}, \tilde{Y})$ is also very close to 0. Neither GND_1, nor GND_2 is as low in cases when $X \not\succeq Y$. So, the empirical estimation of gamma distribution approximates whether stochastic dominance holds well.

What was noted in the description of Table 2 holds as well. GND_1 is always lower than GND_2. Both GND_1 and GND_2 increase with increasing k_Y and θ_Y, and they increase more significantly with increasing θ_Y.

4 Conclusion

We have introduced a procedure that allows the estimation of the general measure of stochastic non-dominance between two random variables. Its application to log-normal and gamma distribution was shown. It seems that the approximations of these distributions have been quite good. The estimated measures of non-dominance were close to zero when we expected them to be. When the estimated general measure of stochastic non-dominance between variables with gamma distribution was compared with the exactly computed measure of non-dominance between corresponding exponential distributions, the approximations were close to the actual values, too. How the estimated $\mathrm{GND}_n(X, Y)$ evolved with the changes in the parameters of Y could be explained as well.

This procedure could be used to approximate the general measures of non-dominance between other distributions with complicated quantile functions. In addition, computing the general measure of non-dominance can be more

complicated than computing the specific measure of non-dominance even for distributions for which the specific measure of non-dominance can be computed quite easily such as the normal distribution. The approximation procedure could therefore be helpful in these cases as well.

This concept could be extended to higher orders of stochastic dominance such as the third-order stochastic dominance studied for example in [14] or decreasing absolute risk aversion stochastic dominance studied for example in [13]. The measures of stochastic non-dominance can be applied also in optimization problems where they may serve as a relaxation of the traditional stochastic dominance constraints. These problems can be formulated also in a multi-stage setting such as in [19]. It can be applied also in dynamic programming, which was recently studied in [7]. Incorporating the endogenous randomness in these programs is another possible extension, which was recently studied in for example [5], [6].

Acknowledgements

The research was supported by the Czech Science Foundation (Grant 19-28231X) and GAUK (grant n. 190123).

References

[1] Ali, M. M. (1975). Stochastic dominance and portfolio analysis. *Journal of Financial Economics*, 2(2),205–229.

[2] Hadar, J. & Russell, W.R. (1969). Rules for ordering uncertain prospects. *American Economic Review*, 59(1),25—34.

[3] Junová, J. & Kopa, M. (2023). Measures of stochastic non-dominance in portfolio optimization. *Submitted to European Journal of Operational Research*.

[4] Kopa, M., Kabašinskas, A., & Šutiene, K. (2021). A stochastic dominance approach to pension-fund selection. *IMA Journal of Management Mathematics*, 33(1),139—160.

[5] Kopa, M. & Rusý, T. (2021). A decision-dependent randomness stochastic program for asset–liability management model with a pricing decision. *Annals of Operations Research*, 299,241—271.

[6] Kopa, M. & Rusý, T. (2023). Robustness of stochastic programs with endogenous randomness via contamination. *European Journal of Operational Research*, 305(3),1259—1272.

[7] Kopa, M. & Šmíd, M. (2023). Contractivity of bellman operator in risk averse dynamic programming with infinite horizon. *Operations Research Letters*, 51(2),133—136.

[8] Levy, H. (2006). *Stochastic Dominance: Investment Decision Making under Uncertainty*. New York: Springer.

[9] Liu, J, Chen, Z. & Consigli, G.(2021). Interval-based stochastic dominance: theoretical framework and application to portfolio choices. *Annals of Operations Research*, 307,329—361.

[10] Moriggia, V., Kopa, M. & Vitali, S. (2019). Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega*, 87,127-–141.

[11] Okagbue, H., Adamu, M.O. & Anake, T.A. (2020). Approximations for the inverse cumulative distribution function of the gamma distribution used in wireless communication. *Heliyon*, 6(11),e05523.

[12] Pflug, G.Ch. & Pichler, A. (2014). Multistage Stochastic Optimization. *Springer*, Cham.

[13] Post, T., Fang, Y. & Kopa, M. (2015). Linear tests for decreasing absolute risk aversion stochastic dominance. *Management Science*, 61(7),1615—1629.

[14] Post, T. & Kopa, M. (2017). Portfolio choice based on third-degree stochastic dominance. *Management Science*, 63(10),3381—3392.

[15] Quirk, J. & Saposnik, R. (1962). Admissibility and measurable utility functions. *Review of Economic Studies*, 29(2),140—146.

[16] Rothschild, M. & Stiglitz, J. (1970). Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3),225—243.

[17] Rothschild, M. & Stiglitz, J. (1971). Increasing risk ii: Its economic consequences. *Journal of Economic Theory*, 3(1),66—84.

[18] Vitali, S. & Moriggia, V. (2021). Pension fund management with investment certificates and stochastic dominance. *Annals of Operations Research*, 229,273—292.

[19] Zapletal, F., Šmíd, M. & Kopa, M. (2020). Multi-stage emissions management of a steel company. *Annals of Operations Research*, 292,735—751.

Bilevel Models in Portfolio Selection Problems

Monika Kaľatová[1]

Abstract. This paper deals with bilevel optimization problem in which we want to minimize transaction costs which are proportional and we want to have our portfolio mean-risk efficient. As a measure of risk we consider Conditional Value-at-Risk. The bilevel optimization problem is presented, reformulations of the problem according to the probability distribution are described as well. The goal of the paper is to present reduction of bilevel optimization problem to a single level optimization problem and to show equivalence between these two types of reduction.

Keywords: bilevel optimization, portfolio optimization, Conditional Value-at-Risk

JEL Classification: C44, G11
AMS Classification: 90C15, 91G10

1 Introduction

Portfolio selection problems are widely used class of problems, since it is the most important type of optimization problems in quantitative finance. Portfolio selection problems has got a lot of attention since Markowitz presented its mean-risk efficient portfolio in [6] and [7]. Since then another risk measures were presented with a focus on coherent risk measures. Conditional Value-at-Risk is often used because of its coherence and computational tractability for discrete distributions. The seminal papers for work with Conditional Value-at-Risk are [10] and [11]. Many more risk measures are stated in [12], but we will focus on mean-Conditional Value-at-Risk objective function in connection with bilevel optimization models.

Bilevel optimization models form a special type of optimization problems, which have two levels. The feasibility set in the upper level problem contains optimal solutions of another optimization problem, which is called the lower level problem. In other words, optimal solutions of the lower level problem become feasible solutions in the upper level problem. We assume that we hold a mean-Conditional Value-at-Risk inefficient portfolio and we want to rebalance it to an efficient one with the minimal transaction costs. Bilevel optimization problems give us an opportunity to solve our problem, because we want to find efficient portfolio and minimize proportional transaction costs at the same time.

This paper is focused on formulating the bilevel optimization problem and its reformulation for different types of probability distributions. The main goal of the paper is to reduce bilevel optimization problem to a single level optimization problem. All theory of bilevel optimization problems in this paper is based on [1].

The rest of the paper is organized as follows. Section 2 presents notation used in the paper and the bilevel optimization problem. Section 3 contains reformulations of the bilevel optimization problem according to the probability distributions. The main part of the paper is in Section 4 which presents reduction of bilevel optimization problem to a single level optimization problem. For this purpose, the primal Karush-Kuhn-Tucker (KKT) transformation and classical Karush-Kuhn-Tucker (KKT) transformation are used.

2 Notation and Formulation

In this section we introduce the bilevel optimization problem and some notation.

2.1 The Bilevel Optimization Problem

Consider the bilevel optimization problem

$$\begin{aligned} &\min_{x,\lambda} ||x - y||_1 \\ &\text{s.t. } \lambda \in \Lambda, \\ &\quad x \in \Psi(\lambda) := \arg\max_{z \in X} (1-\lambda)\mathbb{E}[r^T z] - \lambda \mathrm{CVaR}_\alpha(-r^T z), \end{aligned} \tag{1}$$

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

where the parameter $y \in \mathbb{R}^n$ represents the weights from previous period of investment. The risk aversion parameter λ belongs to the set $\Lambda = [0, 1]$. From the set for portfolio weights, represented by $\mathcal{X} = \{x \in \mathbb{R}^n : \mathbf{1}^T x = 1, x \geq \mathbf{0}\}$, we can see that short selling is not allowed. The reason is that the lower level problem can be then unbounded. The confidence level is represented by parameter $\alpha \in (0, 1)$. Hereafter, $\mathbf{0}$ denotes the null vector of appropriate dimension and $\mathbf{1} = (1, \ldots, 1)^T$ denotes the vector of ones in an appropriate dimension.

Randomness in the bilevel optimization problem (1) is represented by random vector $r = (r_1, \ldots, r_n) \in \mathbb{R}^n$, which consists of returns of underlying assets. It means that the random variable $r_i, i \in \{1, \ldots, n\}$ is the return of the i-th asset. Thus the random return of our portfolio is then defined as $r^T x$, where x are the portfolio weights. Naturally, the random loss is then defined as $-r^T x$. CVaR_α in the bilevel optimization problem (1) denotes Conditional Value-at-Risk at the confidence level α. Since we will work with loss in most of the cases, confidence level α will be chosen close to 1, e.g. $\alpha = 0.95, \alpha = 0.99$.

The interpretation of the bilevel optimization problem (1) is the following. Consider benchmark portfolio which is inefficient with respect to mean and Conditional Value-at-Risk. The investor determines various risk aversion parameters λ in the upper level. Then, for every risk aversion parameter, the lower level problem is calculated, so the set of mean-risk efficient portfolios is obtained. This set of efficient portfolios forms the set of feasible solutions in the upper level problem. Finally, the upper level problem is solved. The resulting portfolio minimizes the investor's transaction costs (for moving from the benchmark portfolio to an efficient one) and is mean-risk efficient as well.

3 Reformulations of Bilevel Optimization Problem

This section is focused on the reformulation of the bilevel optimization problem according to the probability distributions. Firstly, it makes sense to state conditions under which the problem (1) is well-posed.

3.1 Well-posed Problem

The bilevel optimization problem (1) is well-posed in a case that the set $\Psi(\lambda)$ is a singleton for some λ. Then the bilevel problem is optimistic and pessimistic as well. Hence, we need to specify conditions under which the lower level objective function is strictly concave. In this case, the solution to the lower level problem is unambiguous and the investor can take a solution within efficient portfolios which minimizes the upper level objective function.

From the proof of convexity of Conditional Value-at-Risk we need to state conditions under which the following inequality holds strictly almost surely

$$\max\{0, -r^T(x_1 + x_2)\} \leq \max\{0, -r^T x_1\} + \max\{0, -r^T x_2\}, \tag{2}$$

where $x_1, x_2 \in \mathcal{X}$. Using the formula $\max\{a, b\} = 0.5(a + b + |a - b|)$, we obtain that it suffices to establish conditions under which the following equality holds almost surely

$$|r^T(x_1 + x_2)| = |r^T x_1| + |r^T x_2| \tag{3}$$

and these conditions can not hold since we want strict inequality in (2). Obviously, returns of our portfolios for x_1 and x_2 should be nonzero and $x_1 \neq x_2$ ($x_1 \neq \delta x_2$ for $\delta > 0, \delta \neq 1$ holds for all feasible $x_1, x_2 \in \mathcal{X}$). To conclude, for two different weights $x_1 \in \mathcal{X}$ and $x_2 \in \mathcal{X}$, for which the portfolio return is nonzero, the inequality (2) holds strictly and therefore the lower level objective function is strictly concave.

Additionally, we state that the set $\Psi(\lambda)$ can not be empty for every parameter λ, because the set of feasible solution in lower level problem $\mathcal{X}$ is compact and the objective function in lower level problem is continuous, hence we can always find optimal solution for the lower level problem.

3.2 Reformulations

The reformulation of the upper level objective function is obvious since the absolute value function reformulation is known from the theory of single level linear programming. In order to reformulate the lower level program it is necessary to add the assumption for the probability distribution of the random vector r.

Discrete Probability Distribution

Now, we will assume that the probability distribution of the random vector is supported by S atoms. Therefore we assume that we know S scenarios of the random vector r $\varrho^s, s = 1, \ldots, S$ with probabilities $p^s, \forall s \in \{1, \ldots, S\}$.

Using the reformulation for the Conditional Value-at-Risk with scenarios we get the following reformulation of the bilevel optimization problem (1)

$$\begin{aligned}
&\min_{c_i,x,\lambda} \sum_{i=1}^{n} c_i \qquad (4)\\
&\text{s.t. } x_i - y_i \le c_i, \forall i \in \{1,\dots,n\},\\
&\quad -x_i + y_i \le c_i, \forall i \in \{1,\dots,n\},\\
&\quad \lambda \in \Lambda,\\
&\quad (a^*, b^{1^*}, \dots, b^{S^*}, x) \in \Psi(\lambda) := \arg\max_{a,b^s,z} (1-\lambda) \sum_{s=1}^{S} p^s \varrho^{s^T} z - \lambda\Big(a + \frac{1}{(1-\alpha)} \sum_{s=1}^{S} p^s b^s\Big)\\
&\qquad\qquad \text{s.t. } z \in \mathcal{X},\\
&\qquad\qquad\quad b^s \ge -\varrho^{s^T} z - a, \forall s \in \{1,\dots,S\},\\
&\qquad\qquad\quad b^s \ge 0, \forall s \in \{1,\dots,S\}.
\end{aligned}$$

Note that both the lower level problem and the upper level problem in (4) are linear optimization problems. The reformulation (4) of problem (1) for discrete probability distribution will be called the bilevel optimization problem with discrete distribution.

Elliptical Distributions

We will work with elliptical distributions, since they have the property

$$r \sim E(\mu, \Sigma, \phi) \implies r^T x \sim E(\mu^T x, x^T \Sigma x, \phi),$$

where μ is the location vector, Σ is the dispersion matrix and ϕ is the characteristic generator of the distribution. We state normal distribution, generalized Student's t-distribution, the Laplace distribution and the logistic distribution, because these distributions are defined by two parameters (location and scale), as stated in [2]. What's more, the formulas for Conditional Value-at-Risk for these probability distributions can be found in [2]. The general formulation Conditional Value-at-Risk under the assumption of elliptical distributions takes the form

$$-\mu^T z + C\sqrt{z^T \Sigma z},$$

where C is constant corresponding to the given elliptical distribution.

The general reformulation of the bilevel optimization problem for elliptical distributions takes the form

$$\begin{aligned}
&\min_{c_i,x,\lambda} \sum_{i=1}^{n} c_i \qquad (5)\\
&\text{s.t. } x_i - y_i \le c_i, \forall i \in \{1,\dots,n\},\\
&\quad -x_i + y_i \le c_i, \forall i \in \{1,\dots,n\},\\
&\quad \lambda \in \Lambda,\\
&\quad x \in \Psi(\lambda) := \arg\max_{z \in \mathcal{X}} \mu^T z - \lambda C\sqrt{z^T \Sigma z},
\end{aligned}$$

and this formulation differs only by the constant C for various types of elliptical distributions. Constant C is positive for all elliptical distributions and for our values of α as well. Note that the lower level problem in (5) is convex, since the set $\mathcal{X}$ is convex, the objective function is concave and we have the maximization problem. Hereafter, the problem (5) will be called the bilevel optimization problem with elliptical distribution.

Firstly, let us assume that the random vector has n-dimensional normal distribution $r \sim N_n(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$ is the expected value and $\Sigma \in \mathbb{R}^{n\times n}$ is the covariance matrix. For this distribution we have the constant

$$C = \frac{e^{\{\frac{-q_\alpha^2}{2}\}}}{\sqrt{2\pi}(1-\alpha)},$$

where q_α is the α-quantile of $N(0,1)$ distribution.

Secondly, we will assume that the random vector has the n-dimensional generalized Student's t-distribution with $\nu > 3$ degrees of freedom, location parameter μ and Σ as a scale parameter. For this distribution, the constant C takes the form

$$C = \frac{\Gamma(\frac{\nu-1}{2})\sqrt{\nu}}{\Gamma(\frac{\nu-2}{2})(1-\alpha)(\nu-2)\sqrt{\pi}}\left(1+\frac{t^2_{\alpha,\nu}}{\nu}\right)^{-\frac{\nu-1}{2}}.$$

Now let the random vector r have Laplace distribution with μ and Σ parameters of location and scale, respectively. Again, in the reformulation of the bilevel optimization problem (5) the constant C takes the form

$$C = 1 - \log(2\alpha).$$

The last probability distribution is logistic distribution with μ, Σ parameters of location and scale. In the reformulation of the bilevel optimization problem (5) we have constant

$$C = \log \frac{(1-\alpha)^{1-\frac{1}{\alpha}}}{\alpha}.$$

4 Reduction of Bilevel Optimization Problem to a Single Level Optimization Problem

If we want to derive the necessary and sufficient optimality conditions for bilevel optimization problems, the transformation of bilevel optimization problem into a single level optimization problem is usual.

4.1 Primal KKT Transformation

This transformation can be done only in the case when the lower level problem is convex since we want to ensure that the upper level feasibility set will not contain the local optimal solutions or stationary points. This assumption is true in both of our cases: bilevel optimization problem with elliptical distribution (5) and bilevel optimization problem with discrete distribution (4) (linear problem is also convex).

For every risk aversion parameter λ let

$$X(\lambda) := \{x : g(x,\lambda) \leq 0\} = \{x : \mathbf{1}^T x - 1 \leq 0, 1 - \mathbf{1}^T x \leq 0, -x \leq \mathbf{0}\}$$

denote the set of feasible solutions of the lower level problem. Since the constraint function g does not depend on the value of parameter λ, the set $X(\lambda)$ is convex for every value of the risk averse parameter λ. What's more, $X(\lambda) = \mathcal{X}$ for all parameters λ, where $\mathcal{X}$ is defined in bilevel optimization problem (1) and this is the reason why we will work with $\mathcal{X}$ instead of $X(\lambda)$. The objective function of the lower level problem in bilevel optimization problem with elliptical distribution (5) is concave (Conditional Value-at-Risk is convex) and we have the maximization problem, so the assumption for the objective function holds. In our case $T = \mathbb{R}^n$, which is obviously convex and the last assumption for the primal KKT transformation is fulfilled.

Then we get

$$x \in \Psi(\lambda) \Leftrightarrow 0 \in \partial_x f(x,\lambda) + N_{\mathcal{X}\cap T}(x), \tag{6}$$

where $f(x,\lambda)$ is the objective function in the lower level problem of bilevel optimization problem with elliptical distribution (5) (now we use the notation x instead of z) and $\partial_x f(x,\lambda)$ denotes the subdifferential of function f. Since the function $f(\cdot,\lambda)$ is concave in x for every value of parameter λ, then

$$\partial_x f(x,\lambda) = \{-\nabla_x(-f(x,\lambda))\} = \{\nabla_x f(x,\lambda)\}.$$

Immediately we can see that $\mathcal{X} \cap T = \mathcal{X}$, so the normal cone $N_{\mathcal{X}\cap T}(x)$ is defined as

$$N_{\mathcal{X}}(x) := \{d \in \mathbb{R}^n : d^T(z-x) \leq 0 \; \forall z \in \mathcal{X}\}.$$

Finally, we obtain a single level problem equivalent to bilevel optimization problem with elliptical distribution (5),

where C is constant for type of elliptical distribution

$$\begin{aligned}\min_{c_i,x,\lambda,d}\ & \mathbf{1}^T c \\ \text{s.t. } & x_i - y_i \le c_i, \forall i \in \{1,\dots,n\}, \\ & -x_i + y_i \le c_i, \forall i \in \{1,\dots,n\}, \\ & \lambda \in \Lambda, \\ & \mu - \frac{\lambda C}{2\sqrt{x^T\Sigma x}}\Sigma x + d = \mathbf{0}, \\ & d \in N_{\mathcal{X}}(x), \\ & x \in \mathcal{X}.\end{aligned} \tag{7}$$

4.2 Classical KKT Transformation

Let us check the assumptions for the classical KKT transformation. As in primal KKT transformation we have $T = \mathbb{R}^n$ what is a convex set and the function $x \mapsto g(x,\lambda)$ is convex for each fixed risk averse parameter λ.

It is obvious that linearity constraint qualification holds for the lower level problem of bilevel optimization problem with elliptical distribution (5), which implies that Kuhn-Tucker's constraint qualification holds. We want to hold some constraint qualification in order to obtain any KKT point for the lower level problem. Since we have convex problem in the lower level it is sufficient to show that linear independence constraint qualification is satisfied. Gradient for equality constraint $h(x,\lambda) = \mathbf{1}^T x - 1$ is equal to $\nabla_x h(x) = \mathbf{1}$ and gradient for inequality constraint $g_3(x,\lambda) = -x$ is equal to $\nabla_x g_3(x) = e$, where for e holds $e_i = 0$ if $x_i > 0$ and $e_i = 1$ if $x_i = 0$. Gradients $\mathbf{1}$ and e would be linearly dependent if $e = \mathbf{1}$, but in this case $x = \mathbf{0}$, which is infeasible vector for our problem. Hence the linear independence constraint qualification holds.

Assumptions for the classical KKT transformation are satisfied, so let us derive the KKT conditions for the lower level problem of bilevel optimization problem with elliptical distribution (5). Let $\beta \in \mathbb{R}, \gamma \in \mathbb{R}^n$ be Lagrange multipliers, then we get

1. primal feasibility condition: $x \in \mathcal{X}$,
2. dual feasibility condition: $\frac{\lambda C}{2\sqrt{x^T\Sigma x}}\Sigma x - \mu + \beta\mathbf{1} - \gamma^T x = \mathbf{0}$,
3. complementary slackness condition: $\gamma^T x = 0, \gamma \ge \mathbf{0}$.

Using these KKT conditions we get the second reformulation of the bilevel optimization problem with elliptical distribution (5), where C again denotes constant for type of elliptical distribution

$$\begin{aligned}\min_{c_i,x,\lambda,\beta,\gamma}\ & \mathbf{1}^T c \\ \text{s.t. } & x_i - y_i \le c_i, \forall i \in \{1,\dots,n\}, \\ & -x_i + y_i \le c_i, \forall i \in \{1,\dots,n\}, \\ & \lambda \in \Lambda, \\ & x \in \mathcal{X}, \\ & \frac{\lambda C}{2\sqrt{x^T\Sigma x}}\Sigma x - \mu + \beta\mathbf{1} - \gamma = \mathbf{0}, \\ & \gamma^T x = 0, \\ & \gamma \ge \mathbf{0}.\end{aligned} \tag{8}$$

4.3 Linking Primal KKT Transformation and Classical KKT Transformation

It seems that if some $\hat{x}$ is feasible for the Primal KKT reformulation (7), then it is feasible for the Classical KKT reformulation (8) and vice versa.

Firstly, we will assume that $(c,\hat{x},\lambda,\beta,\gamma)$ is feasible for the Classical KKT reformulation (8). It is sufficient to show that from dual feasibility condition and complementary slackness condition vector $-(\beta\mathbf{1}-\gamma)$ lies in the opposite of normal cone $N_{\mathcal{X}}(\hat{x})$ since the KKT conditions were formulated for the minimization problem and the normal cone was formulated for the maximization problem. It is true, because

$$-(\beta\mathbf{1}-\gamma)^T(z-\hat{x}) = \gamma^T z \ge 0, \forall z \in \mathcal{X},$$

where we used properties of the vector z and the complementary slackness condition.

Secondly, let $(c, \hat{x}, \lambda, d)$ be feasible for the Primal KKT reformulation (7). Every canonical vector e_i is feasible for our lower level problem. Since we know that $d \in N_X(\hat{x})$, it has to hold

$$d_i \leq d^T\hat{x}, \forall i \in \{1, \ldots, n\}.$$

If we want to find Lagrange multipliers we have to find non-positive Lagrange multipliers for the lower level problem, since we pass from the maximization problem to the minimization problem. Let us denote $\gamma = d - (d^T\hat{x})\mathbf{1} \leq 0$. If we want γ to be Lagrange multiplier, we have to verify $0 = \gamma^T\hat{x} = (d - (d^T\hat{x})\mathbf{1})^T\hat{x}$, what obviously holds. It suffices to set $\beta = -d^T\hat{x} \in \mathbb{R}$ and we have found the second Lagrange multiplier. Finally, we obtain the decomposition of the feasible vector d such that $(c, \hat{x}, \lambda, \beta, \gamma)$ is feasible for the Classical KKT reformulation (8).

5 Conclusion

This paper presents bilevel optimization problem with application to portfolio theory (1). We state its reformulations for discrete probability distribution (4) and elliptical distributions (5). The main focus of this paper is on the reduction of bilevel optimization problem in a single level optimization problem and linking two types of transformations used for the reduction.

This paper can be extended in many ways. Firstly, we can reformulate the bilevel optimization problem for other probability distributions. We can apply the bilevel optimization problem with discrete distribution (4) to historical scenarios. Additionally, we can consider portfolio efficiency in different sense, e.g. we can consider stochastic dominance as in [8] and [9]. We can work with multi-stage problems with mean-Conditional Value-at-Risk portfolio efficiency [13] or with dynamic programming [5]. Moreover, we can consider endogenous randomness, see [3] and [4] for more details.

Acknowledgements

This research was supported by Czech Science Foundation [Grant No. 19-28231X] and GAUK [Grant No. 178723].

References

[1] Dempe, S. et al. (2015). *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*. Springer Berlin Heidelberg.

[2] Khokhlov, V. (2016). Conditional Value-at-Risk for Elliptical Distributions. *European Journal of Economics and Management*, 2, 6, 70-79.

[3] Kopa, M. & Rusý, T. (2021). A decision-dependent randomness stochastic program for asset–liability management model with a pricing decision. *Annals of Operations Research*, 299, 1, 241-271.

[4] Kopa, M. & Rusý, T. (2023). Robustness of stochastic programs with endogenous randomness via contamination. *European Journal of Operational Research*, 305, 3, 1259-1272.

[5] Kopa, M. & Šmíd, M. (2023). Contractivity of Bellman operator in risk averse dynamic programming with infinite horizon. *Operations Research Letters*, 51, 133-136.

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

[7] Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.

[8] Post, T. & Kopa, M. (2013). General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research*, 230, 2, 321–332.

[9] Post, T., Fang, Y. & Kopa, M. (2015). Linear Tests for DARA Stochastic Dominance. *Management Science*, 61, 7, 1615-1629.

[10] Rockafellar, R. T. & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 3, 21–41.

[11] Rockafellar, R. T. & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26, 7, 1443–1471.

[12] Szegö, G. (ed.) (2004). *Risk measures for the 21st century*. Wiley.

[13] Zapletal, F., Šmíd, M. & Kopa, M. (2020). Multi-stage emissions management of a steel company. *Annals of Operations Research*, 292, 2, 735-751.

Ambiguity in Stochastic Optimization Problems with Nonlinear Dependence on a Probability Measure via Wasserstein Metric

Vlasta Kaňková [1]

Abstract. Many economic and financial applications lead to deterministic optimization problems depending on a probability measure. It happens very often (in applications) that these problems have to be solved on the data base. Point estimates of an optimal value and estimates of an optimal solutions set can be obtained by this approach. A consistency, a rate of convergence and normal properties, of these estimates, have been discussed (many times) not only under assumptions of independent data corresponding to the distributions with light tails, but also for weak dependent data and the distributions with heavy tails. However, it is also possible to estimate (on the data base) a confidence intervals and bounds for the optimal value and the optimal solutions. To analyze this approach we focus on a special case of static problems depending nonlinearly on the probability measure. Stability results based on the Wasserstein metric and the Valander approach will be employed for the above mentioned analysis.

Keywords: Stochastic optimization problems, static problems, empirical measure, point estimates, interval estimates, nonlinear dependence

JEL classification: C44
AMS classification: 90C15

1 Introduction

To introduce a primary "classical" stochastic static one-objective optimization problem, 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 $(\Omega, \mathcal{S}, P)$; $F(:= F_\xi(z), z \in R^s)$ the distribution function of ξ; P_F, Z_F the probability measure and a support corresponding to F; $X_F \subset X \subset R^n$ a nonempty set generally depending on F; $X \subset R^n$ a nonempty "deterministic" set; E_F an operator of mathematical expectation corresponding to the distribution function F. If $g_0(x, z)$ is a real-valued function defined on $R^n \times R^s$, then a primary classical problem of the stochastic optimization can be (in a rather general setting) introduced in the form:

Find

$$\varphi(F,\, X) = \inf\{\mathsf{E}_F g_0(x, \xi) | x \in X\}. \tag{1}$$

The "deterministic" constraints set X, in the problem (1), has been (from the beginning of the stochastic optimization) often replaced by the set X_F depending on the probability measure (see, e.g., [1] or [12]). Simultaneously, it has been soon recognized that these problems have to be often solved (in applications) on the data base; the probability measure P_F has to be often replaced by an empirical measure P_{F^N}. Consequently, instead of the original Problem (1) the following Empirical Problem often has to be solved:

Find

$$\varphi(F^N,\, X_{F^N}) = \inf\{\mathsf{E}_{F^N} g_0(x, \xi) | x \in X_{F^N}\}. \tag{2}$$

A great effort has been paid to investigate a relationship between an optimal value and an optimal solution of the problem (1) with $X := X_F$, generally, and they point estimates obtained by the problem (2). Works dealing with confidence intervals have begun to appear about year 2000. We can recall here the paper [6],

[1] The Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, kankova@utia.cas.cz

where the Kolmogorov distance has been employed to study an ambiguity. The Kantorovich distance has been employed to get ambiguity results in [11], the ambiguity in chance constrained problems has been investigated in [3]. The optimization problems with a nonlinear dependence on the probability measure have begun more to appear only in the last time (see, e.g., [2], [4], [7] or [8]).

To recall optimization problems with a nonlinear dependence on the probability measure, let $\bar{g}_0(:= \bar{g}_0(x,z,y))$ be a real–valued function defined on $R^n \times R^s \times R^{m_1}$; $h(:= h(x,z)) = (h_1(x,z),\ldots,h_{m_1}(x,z))$ be an m_1–dimensional vector function defined on $R^n \times R^s$. A stochastic static one-objective optimization problem with the nonlinear dependence on the probability measure can be introduced in the form:

Find

$$\bar{\varphi}(F,\, X_F) = \inf\{\mathsf{E}_F \bar{g}_0(x,\xi,\mathsf{E}_F h(x,\xi))|x \in X_F\}, \tag{3}$$

where a nonlinear dependence can appear also in the constraints set X_F. We consider the types of X_F :

$$\begin{aligned} a. \qquad X_F &:= X, \\ b. \qquad X_F &:= \{x \in X : \mathsf{E}_F \bar{g}_i(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)) \le 0,\, i = 1,\, \ldots,\, m\}, \\ &\qquad \text{where} \quad \bar{g}_i(x,\, z,\, y),\, i = 1,\, \ldots,\, m \quad \text{are defined on} \quad R^n \times R^s \times R^{m_1}. \end{aligned} \tag{4}$$

Of course, we consider that all mathematical expectations in (1), (2), (3), (4) exist and they are finite.

To define second order stochastic dominance constraints set let $Y(\xi)$, $g(x,\, \xi)$ be for every $x \in X$ random variables with distribution function $F_{Y(\xi)}$, $F_{g(x,\,\xi)}$. Let, moreover, for every $x \in X$ there exist finite $\mathsf{E}_F g(x,\xi)$, $\mathsf{E}_F Y(\xi)$ and

$$F^2_{g(x,\,\xi)}(u) = \int_{-\infty}^{u} F_{g(x,\,\xi)}(v)dv, \quad F^2_{Y(\xi)}(u) = \int_{-\infty}^{u} F_{Y(\xi)}(v)dv, \quad u \in R^1.$$

Rather general second order stochastic dominance constraints set X_F can be defined by

$$c. \qquad X_F \,=\, \{x \in X :\, F^2_{g(x,\,\xi)}(u) \le F^2_{Y(\xi)}(u) \quad \text{for every} \quad u \in R^1\},$$

or equaivalently by

$$X_F = \{x \in X :\, \mathsf{E}_F(u - g(x,\xi))^+ \,\le\, \mathsf{E}_F(u - Y(\xi))^+ \quad \text{for every} \quad u \in R^1\}. \tag{5}$$

The proof of the last equivalence can be found in [10].

Very often it is necessary (instead of the underlying problem (3) also here) to solve empirical problem

Find

$$\bar{\varphi}(F^N,\, X_{F^N}) = \inf\{\mathsf{E}_{F^N} \bar{g}_0(x,\xi,\mathsf{E}_{F^N} h(x,\xi))|x \in X_{F^N}\}. \tag{6}$$

2 Some Definitions, Assumptions and Auxiliary Assertion

Our analysis of the ambiguity is based on the Wasserstein metric and $\mathcal{L}_1$ distance in R^s . To this end, let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures on R^s and let the system $\mathcal{M}^1_1(R^s)$ be defined by the relation:

$$\mathcal{M}^1_1(R^s) \,=\, \{\nu \in \mathcal{P}(R^s) : \int_{R^s} \|z\|_1 d\nu(z) < \infty\}, \quad \|\cdot\|_1 \quad \text{denotes} \quad \mathcal{L}_1 \quad \text{norm in} \quad R^s.$$

2.1 Definitions and Assumptions

First, we define a system of the assumptions:

A.1 1. $g(x,z)$, $Y(z)$ are for $x \in X$ Lipschitz functions of $z \in R^s$ with the Lipschitz constant L_g (corresponding to the $\mathcal{L}_1$ norm) not depending on x,

A.2 1. $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F,

2. F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^{N}$, $N = 1, 2, \ldots,$

B.1 $P_F \in \mathcal{M}_1^1(R^s)$, there exist $\varepsilon > 0$ and an ε-neighbourhood $X(\varepsilon)$ of X such that

1. $\bar{g}_0(x,\, z,\, y)$ is, for $x \in X(\varepsilon)$, $z \in R^s$, a Lipschitz function of $y \in Y(\varepsilon)$ with the Lipschitz constant L_y; $Y(\varepsilon) = \{y \in R^{m_1} : y = h(x,\, z) \quad \text{for some} \quad x \in X(\varepsilon),\, z \in R^s\}$, $\mathsf{E}_F h(x,\, \xi)$, $\mathsf{E}_{F^N} h(x,\, \xi) \in Y(\varepsilon), \quad \text{for} \quad x \in X(\varepsilon),\ N = 1,\, 2,\, \ldots,$
2. for every $x \in X(\varepsilon)$ and every $y \in Y(\varepsilon)$ there exist finite mathematical expectations $\mathsf{E}_F \bar{g}_0(x,\, \xi,\, y)$, $\mathsf{E}_{F^N} \bar{g}_0(x,\, \xi,\, y)$,
3. $h_j(x,\, z)$, $j = 1,\, \ldots,\, m_1$ are for every $x \in X(\varepsilon)$ Lipschitz functions of $z \in R^s$ with the Lipschitz constants L_h^i (corresponding to the $\mathcal{L}_1$ norm),
4. $\bar{g}_0(x,\, z,\, y)$ is for every $x \in X(\varepsilon)$, $y \in Y(\varepsilon)$ a Lipschitz function of $z \in R^s$ with the Lipschitz constant L_z (corresponding to the $\mathcal{L}_1$ norm),

B.2 $\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi))$, $\mathsf{E}_{F^N} \bar{g}_0(x, \xi, \mathsf{E}_{F^N} h(x, \xi))$, $N = 1,\, \ldots$ are continuous functions on X,

C.1
- $\bar{g}_0(x,\, z,\, y)$ is for every $z \in Z_F$ and $y \in Y(\varepsilon)$ a Lipschitz function of $x \in X$ with the Lipschitz constant L_C not depending on $z \in Z_F$, $y \in Y(\varepsilon)$,
- $h_j(x,\, z)$, $j = 1,\, \ldots,\, m_1$ are for every $z \in Z_F$ Lipschitz functions on X with the Lipschitz constant L_C^h not depending on $z \in Z_F$.

Further, we recall two Definitions and simultaneously define for $\varepsilon \in R^1$ the sets $X_F^{b,\,\varepsilon}$, $X_F^{c,\,\varepsilon}$.

Definition 1. [13] Let $\|\cdot\| = \|\cdot\|_n$denote the Euclidean norm in R^n. If $X^{'}$, $X^{''} \subset R^n$ are two non–empty sets, then the Hausdorff distance of these sets $\Delta[X^{'},\, X^{''}] := \Delta_n[X^{'},\, X^{''}]$ is defined by

$$\Delta_n[X^{'},\, X^{''}] \,=\, \max[\delta_n(X^{'},\, X^{''}),\, \delta_n(X^{''},\, X^{'})], \quad \delta_n(X^{'},\, X^{''}) \,=\, \sup_{x^{'} \in X^{'}} \inf_{x^{''} \in X^{''}} \|x^{'} - x^{''}\|.$$

Definition 2. [13] Let $\hat{h}(x)$ be a real–valued function defined on a nonempty convex set $\mathcal{K} \subset R^n$. $\hat{h}(x)$ is a strongly convex function on $\mathcal{K}$ with a parameter $\bar{\rho} > 0$ if

$$\hat{h}(\lambda x^1 + (1-\lambda)x^2) \,\leq \lambda \hat{h}(x^1) + (1-\lambda)\hat{h}(x^2) - \lambda(1-\lambda)\bar{\rho}\|x^1 - x^2\|_n^2 \quad \text{for every} \quad x^1,\, x^2 \in \mathcal{K},\, \lambda \in \langle 0,\, 1\rangle.$$

$$\begin{aligned} X_F^{b,\,\varepsilon} \;&=\; \{x \in X :\, \mathsf{E}_F \bar{g}_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)) \leq \varepsilon\} \quad \text{in the case of the constraints set} \quad b.) \ \text{with } m = 1, \\ X_F^{c,\,\varepsilon} \;&=\; \{x \in X :\, \mathsf{E}_F(u - g(x,\, \xi))^+ - \mathsf{E}_F(u - Y(\xi))^+ \leq \varepsilon \quad \text{for every} \quad u \in R^1\} \\ &\qquad \text{in the case of the constraints set} \quad c.). \end{aligned} \tag{7}$$

Evidently $X_F = X_F^0 := X_F^{b,\,0}$ in the case b.) with $m = 1$; $X_F = X_F^0 := X_F^{c,\,0}$ in the case set c.).

2.2 Brief Survey of Former Results

Lemma 1. [9] Let $\mathcal{K} \subset R^n$ be a nonempty, convex, compact set, $x_0 = \arg\min_{x \in \mathcal{K}} \hat{h}(x)$, If

1. $\hat{h}(x)$ is a strongly convex continuous function on $\mathcal{K}$ with a parameter $\bar{\rho} > 0$, $\mathcal{K}^{\varepsilon} = \{x \in \mathcal{K} : \hat{h}(x) \leq \varepsilon\}$, $\varepsilon > \hat{h}(x_0)$,
2. for $\varepsilon_1,\ \varepsilon_2 > \hat{h}(x_0), \quad \varepsilon_1 < \varepsilon_2 \quad$ it holds that
   - there exists $x_2 \in \mathcal{K}$ such that $\hat{h}(x_2) = \varepsilon_2$,
   - for every $x_2 \in \mathcal{K}^{\varepsilon_2}$, $\hat{h}(x_2) = \varepsilon_2$, there exists a projection $x_1 := x_1(x_2)$ on $\mathcal{K}^{\varepsilon_1}$, $\hat{h}(x_1) = \varepsilon_1$,
   - $\hat{h}(x(\lambda))$ is an increasing function of λ, $\lambda \in \langle 0,\, 1\rangle, \quad x(\lambda) = \lambda x_2 + (1-\lambda)x_1$,

then $$[\Delta_n[\mathcal{K}^{\varepsilon_1},\, \mathcal{K}^{\varepsilon_2}]]^2 \le \frac{2}{\bar\rho}|\hat h(x_2) - \hat h(x_1)|.$$

If we denote by $F_i(z_i)$, $i = 1, \ldots, s$ one dimensional marginal distributions corresponding to F and by $\mathcal{N}_0$ the set of all natural numbers, then we can recall the following assertion.

Theorem 1. [7] Let $P_F \in \mathcal{M}_1^1(R^s)$ and let, moreover, X be a nonempty compact set, Assumptions A.2, B.1, B.2 be fulfilled, then for every $N \in \mathcal{N}_0$ it holds

$$|\bar\varphi(F,\, X) - \bar\varphi(F^N,\, X)| \le D \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - F_i^N(z_i)|dz_i, \quad \text{where} \quad 0 \le D \le L_y \sum_{j=1}^{m_1} L_h^i + L_z. \tag{8}$$

Further, we little modify the assertions of the paper [9].

Theorem 2. Let $P_F \in \mathcal{M}_1^1(R^s)$, X be nonempty, compact convex sets, X_F be given by the constraint set b.) with $m = 1$. Let, moreover, X_F, X_{F^N}, $N \in \mathcal{N}$, be nonempty compact sets. If

1. $\mathsf{E}_F\bar g_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi))$ is a strongly convex, with a parameter $\bar\rho > 0$, function on X that fulfils the assumptions of Lemma 1 (setting $\mathsf{E}_F\bar g_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)) := \hat h(x),\ X := \mathcal{K}$),
2. $x_0 = \arg\min\limits_{x\in X} \mathsf{E}_F\bar g_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)), \quad \varepsilon_1 > \mathsf{E}_F\bar g_1(x_0,\, \xi,\, \mathsf{E}_F h(x_0,\, \xi))$,
3. $X_F := X_F^{\varepsilon_1} = \{x \in X :\, \mathsf{E}_F\bar g_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)) \le \varepsilon_1\}$,
4. - Assumption A.2 is fulfilled,
   - $\bar g_1$ fulfils Assumptions B.1, B.2, (setting $\bar g_1 := \bar g_0$), $\bar g_0$ fulfils Assumptions B.1, B.2, C.1,
5. there exists $\bar\varepsilon_0 > 0$, $\bar\varepsilon_0 := \bar\varepsilon_0(\varepsilon_1)$ such that $X_F^{\varepsilon_1 - \varepsilon_0}$ is for $0 < \varepsilon_0 < \bar\varepsilon_0$ a nonempty compact set,

then for $N \in \mathcal{N}_0$, $\hat C = L_y \sum\limits_{j=1}^{m_1} L_h^i + L_z$ fulfilling the relations

$$\hat C[\sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) \,-\, F_i^N(z_i)|dz_i] \,\le\, \bar\varepsilon_0, \quad \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) \,-\, F_i^N(z_i)|dz_i \,\le\, 1, \tag{9}$$

the next assertion is valid

$$|\bar\varphi(F,\, X_F) - \bar\varphi(F^N,\, X_{F^N})| \le D \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) \,-\, F_i^N(z_i)|dz_i, \quad D = \hat C + 2[\max[L_C,\, m_1 L_y\, L_C^h][\frac{4}{\bar\rho}\hat C 10]^{1/2}. \tag{10}$$

Theorem 3. Let $P_F \in \mathcal{M}_1^1(R^s)$, $X \subset R^n$ be a nonempty compact set, X_F correspond to the constraint set b.) with $m = 1$, X_F, X_{F^N}, $N = 1, 2, \ldots$ be nonempty compact sets, If

1. - Assumption A.2 is fulfilled,
   - $\bar g_1$ fulfils Assumptions B.1, B.2, (setting $\bar g_1 := \bar g_0$), $\bar g_0$ fulfils Assumptions B.1, B.2, C.1,
2. there exists $\bar\varepsilon_0 > 0$ such that X_F^{ε} (defined by the relation (7)) are nonempty compact sets for every $\varepsilon \in \langle -\bar\varepsilon_0,\, \bar\varepsilon_0\rangle$ and, moreover, there exists a constant $\bar C > 0$ such that

$$\Delta_n[X_F^{\varepsilon},\ X_F^{\varepsilon'}] \,\le \bar C|\varepsilon - \varepsilon'| \quad \text{for} \quad \varepsilon, \varepsilon' \in \langle -\bar\varepsilon_0,\, \bar\varepsilon_0\rangle,$$

then for $N \in \mathcal{N}_0$ fulfilling inequality $\hat C \sum\limits_{i=1}^{s} \int\limits_{-\infty}^{\infty} |F_i(z_i) \,-\, F_i^N(z_i)|dz_i] \le \bar\varepsilon_0$ it holds that

$$|\bar\varphi(F,\, X_F) - \bar\varphi(F^N,\, X_{F^N})| \,\le\, D\sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) \,-\, F_i^N(z_i)|dz_i, \quad D \,=\, \hat C[1 + 2\max[L_C,\, 2m_1 L_y L_C^h]10\bar C]. \tag{11}$$

If the assumption 2 holds for every $\bar\varepsilon_0 \in R^1$, then the inequality (11) holds also for all $N \in \mathcal{N}_0$.

Theorem 4. Let $P_F \in \mathcal{M}_1^1(R^s)$, $X \subset R^n$ be a nonempty compact set, X_F correspond to the constraints set c.), Assumption A.1, A.2 be fulfilled and let X_F, X_{F^N}, $N = 1, 2 \ldots$ be nonempty compact sets. If

1. $\bar{g}_0$ fulfils Assumptions B.1, B.2, C.1,
2. there exists $\bar{\varepsilon}_0 > 0$ such that X_F^{ε} (defined by the relation (7)) are nonempty compact sets for every $\varepsilon \in \langle -\bar{\varepsilon}_0,\, \bar{\varepsilon}_0 \rangle$ and, moreover, there exists a constant $\bar{C} > 0$ such that

$$\Delta_n[X_F^{\varepsilon},\ X_F^{\varepsilon'}] \ \leq \bar{C}|\varepsilon - \varepsilon'| \quad \text{for} \quad \varepsilon, \varepsilon' \in \langle -\bar{\varepsilon}_0,\, \bar{\varepsilon}_0 \rangle,$$

then for $N \in \mathcal{N}_0$ fulfilling inequality $2L_g \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - F_i^N(z_i)|dz_i] \leq \bar{\varepsilon}_0$ it holds that

$$|\bar{\varphi}(F,\, X_F) - \bar{\varphi}(F^N,\, X_{F^N})| \ \leq \ D \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - F_i^N(z_i)|dz_i, \quad D \ = \ \hat{C}[1 + 2\max[L_C,\, m_1 L_y L_C^h]20\bar{C}L_g \tag{12}$$

If the assumption 2 holds for every $\bar{\varepsilon}_0 \in R^1$, then the inequality (12) holds also for all $N \in \mathcal{N}_0$ [9].

Further, we recall Kolmogorov's limit Theorem. To this end we consider $s = 1$.

Proposition 1. [5] Let $s = 1$. If the probability measure corresponding to $F(z)$ is absolutely continuous with respect to the Lebesgue measure on R^1, Assumption A.2 is fulfilled, then

$$\begin{aligned} \lim_{N\to\infty} P\{\omega :\, (N)^{1/2} \sup_z |F(z) - F^N(z)| \leq t\} \quad &= \quad \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2t^2} \quad \text{for} \quad t > 0, \\ &= 0 \qquad\qquad\qquad\quad \text{for} \quad t \leq 0. \end{aligned} \tag{13}$$

Evidently, in this case $K^N(t) = P\{\omega : \sup_z |F(z) - F^N(z)| \leq t\}$ for $t \in R^1$, $N = 1, 2 \ldots$, is the distribution function of the random value $\sup_z |F(z) - F^N(z)|$, its quantils for $N \geq 100$ can be approximated employing the relation (13) (see, e.g., [5]).

3 Ambiguity Analysis

To analyze ambiguity, first, we introduce the following assumption:

C.2
- P_{F_i}, $i = 1, \ldots, s$ are absolutely continuous w.r.t. Lebesgue measure on R^1;
- $Z_{F_i} = \langle a_i,\, b_i \rangle,\ \langle a_i,\, b_i \rangle \subset \langle a,\ b \rangle$ for some $a_i,\, b_i,\, a,\, b \in R^1,\, a_i \leq b_i,\, a \leq b,\, i = 1, \ldots,\ s,$

and we define $k_N(\alpha)$ for $\alpha \in (0,\, 1)$, $N \in \mathcal{N}$ by

$$P\{\omega \in \Omega : \sup_{z_i} |F_i(z_i) - F_i^N(z_i)| \geq k_N^i(\alpha)\} \leq \alpha, i = 1, 2, \ldots, s, \qquad k_N(\alpha) = \sup_i k_N^i(\alpha).$$

Consequently if $1 - s\alpha > 0$, then according to C.2 and Theorems of Section 2 (we can obtain successively (for the corresponding constant D) that

$$\begin{aligned} P\{\omega \in \Omega : \bigcup_{i=1}^{s} [\sup_{z_i} |F_i(z_i) - F_i^N(z_i)| \geq k_N(\alpha)]\} \quad &\leq \quad s\alpha, \qquad\qquad N \in \mathcal{N}_0, \\ P\{\omega \in \Omega : \bigcap_{i=1}^{s} [\sup_{z_i} |F_i(z_i) - F_i^N(z_i)| < k_N(\alpha)]\} \quad &> \quad 1 - s\alpha, \qquad\qquad N \in \mathcal{N}_0, \\ P\{\omega : D \sum_{i=1}^{s} \int_{a_i}^{b_i} |F_i(z_i) - F_i^N(z_i)|dz_i < dsDk_N(\alpha)\} \quad &> \quad 1 - s\alpha, \quad d = b - a, \quad N \in \mathcal{N}_0, \\ .P\{\omega : |\bar{\varphi}(F,\, X) - \bar{\varphi}(F^N,\, X_{F^N})| \ < \ dsDk_N(\alpha)\} \quad &> \quad 1 - s\alpha, \qquad\qquad N \in \mathcal{N}_0. \end{aligned}$$

Further, if we denote by $\mathcal{F}$ a system of all $s-$ dimensional distribution functions F with one dimensional marginal distribution functions F_i, $i = 1, \ldots, s$ fulfilling the assumption C.2, then

$$F \in \mathcal{F} \Rightarrow P\{\omega : |\bar{\varphi}(F,\, X) - \bar{\varphi}(F^N,\, X_{F^N})| < dsDk_N(\alpha)\} > 1 - s\alpha, \quad N \in \mathcal{N}_0. \tag{14}$$

Theorem 5. Let $X \subset R^n$ be a nonempty compact set, assumption C.2 be fulfilled, $\alpha \in (0,\, 1)$, $1 - s\alpha > 0$, $F \in \mathcal{F},\ N \in \mathcal{N}_0$ fulfil the corresponding Theorem of Section 2. If

1. the assumptions of Theorem 1 are fulfilled, $D = L_y \sum_{j=1}^{m_j} L_h^i + L_z$, then

$$F \in \mathcal{F} \Rightarrow P\{\omega : |\bar{\varphi}(F, X) - \bar{\varphi}(F^N, X_{F^N})| < dDsk_N(\alpha)\} > 1 - s\alpha,\ N = 1, 2, \ldots, \tag{15}$$

2. the assumptions of Theorem 2 are fulfilled, $D = \hat{C}[1 + 2[\max[L_C, m_1 L_y L_C^h][\frac{2}{\bar{\rho}}10]^{1/2}]$, then

$$F \in \mathcal{F} \Rightarrow P\{\omega :\ |\bar{\varphi}(F, X_F) - \bar{\varphi}(F^N, X_{F^N})| < dDsk_N(\alpha) > 1 - s\alpha\} \tag{16}$$

3. the assumptions of Theorem 3 are fulfilled, $D = \hat{C} + 2\max[L_C, m_1 L_y L_C^h]10\hat{C}\bar{C}$,

$$F \in \mathcal{F} \Rightarrow P\{\omega : |\bar{\varphi}(F, X_F) - \bar{\varphi}(F^N, X_{F^N})| < dDsk_N(\alpha)\} > 1 - s\alpha), \tag{17}$$

4. the assumptions of Theorem 4 are fulfilled, $D = \hat{C} + \max[L_C, m_1 L_y L_C^h]\bar{C}20L_g$. then

$$F \in \mathcal{F} \Rightarrow P\{\omega :\ |\bar{\varphi}(F, X_F) - \bar{\varphi}(F^N, X_{F^N})| < dDsk_N(\alpha)\}\} > 1 - s\alpha, \tag{18}$$

Consequently, under the assumptions of Theorem 5, we get.

$$P\{\omega : \bar{\varphi}(F^N,\ X_{F^N}) - dDsk_N(\alpha)\ < \bar{\varphi}(X, X_F)\ < \bar{\varphi}(F^N,\ X_{F^N}) + dDsk_N(\alpha)\} > 1 - s\alpha \tag{19}$$

with constant D determined by the corresponding results given in Theorem 5.

4 Conclusion

The contribution is focused on a special type of the stochastic optimization problems in which dependence on the probability measure is not linear. This type of problems corresponds to real-life situations rather often and it has been investigated in [8], [9]. This contribution tries to employ there achieved results to investigate ambiguity properties.

Acknowledgement

This work was supported by the Czech Science Foundation under grant 18-02739S.

References

[1] Birge, J.R. & Louveaux, F. (1999). *Introduction to Stochastic Programming.* Springer: Berlin.
[2] Dentcheva, D., Penev, S. & Ruszczynski, A. (2014). Statistical estimation of composite risk functionals and risk optimization problem, *Conwelt University Library.*
[3] Erdogan, E. & Iyengar, G. (2005). Ambigious chance constrained and robust optimization, *SPEPS.*
[4] Ermoliev, Yu. & Norkin, V. (2013). Sample average approximation method for compound stochastic optimization problems. *SIAM J. Optimization*, 23 (4), 2231–2263.
[5] Janko, J. (1958). *Statistical Tables (in Czech).* Czechoslovak Academy of Sciences (in Czech).
[6] Kaňková, V. (1996). A note on interval estimates in stochastic optimization. *Bulletin of the Czech Econometric Society* 5, 63–79.
[7] Kaňková, V. & Houda, M. (2015). Thin and heavy tails in stochastic programming. *Kybernetika*, 51(3), 433–456.
[8] Kaňková, V. (2020). A note on stochastic optimization problems with nonlinear dependence on a probability measure. *Proceedings of the 38th International Conference Mathematical Methods in Economics 2020* S. Kapounek and H. Vránová, eds.), Mendel University in Brno, Faculty of Business and Economics, Brno, 247–252 (2020)
[9] Kaňková, V. (2022) Stochastic optimization problems with nonlinear dependence on a probability measure via Wasserstein metric. *Journal of Global Optimization, submitted.*
[10] Ogryczak, W. & Rusczynski, A. (1999). From the stochastic dominance to mean–risk models: semideviations as risk measure. *European J. Oper. Res.*, 116, 33–50.
[11] Pflug, G.Ch., Pichler, A. & Wozabal, D. (2012). The 1/N investment strategy to optimal under high mode ambiguity. *Journal of Banking & Finance* (36) 410–417.
[12] Prékopa, A. (1995). *Stochastic Programming.* Akadémia Kiadó, Budapest and Kluwer: Dordrecht 1995.
[13] Rockafellar, R. & Wets, R.J.B. (1983). *Variational Analysis.* Berlin: Springer.

Heterogeneous Effects of Financial Uncertainty: Evidence from Global Financial Crisis

Svatopluk Kapounek[1], Roman Horvath[2]

Abstract. We measure financial uncertainty employing textual analysis of newspaper articles in the period from 1985 to 2017. We show boosted effects of financial uncertainty shocks on the economic activity after the financial crisis. However, we identify lower negative effects of financial uncertainty on the interest rates after the crisis period. In addition, we provide analysis of financial uncertainty effects on several financial indicators and confirm negative effects of financial uncertainty on house prices after the financial crisis and positive effects on several spreads before the crisis periods.

Keywords: financial market, uncertainty, news media, global financial crisis

JEL Classification: D80, G10, G20
AMS Classification: 62P20

1 Introduction

The economic policy uncertainty is defined as the uncertain effect of the economic policy changes related to political costs and benefits [23]. Such an uncertain effects of the economic policy are related to the global financial crisis in 2007 [7, 8], especially due to the asymmetric information and new regulatory environment around the world [20, 15]. The government's future actions and the economic policy effects are uncertain because the effects of crisis are heterogeneous. Thus, there are not identical responses of the economic policy on the shocks, there are not identical effects of economic policy changes on the real economy and financial markets.

There is wide range of literature on the heterogeneous effects of global financial crisis. E.g. in terms of government spending [3, 4], monetary policy and bank lending changes [6, 10, 15, 24], consequent economic and investment activity [2, 14, 18, 19]. Moreover, Caggiano et al. [12] shows deepened impact of economic policy uncertainty during the recessions.

However, the global financial crisis in 2007 is caused by the financial shocks associated with increasing uncertainty [9]. Therefore, we contribute with employing uncertainty at the financial markets and construct the new index of financial uncertainty. Following the logic above, we define our first hypothesis:

H1: the financial uncertainty effects on the economic activity increased after the global financial crisis in 2007.

In addition, Aikman et al. [1] show that financial uncertainty is more related to the financial variables, especially spreads and lending conditions. We suppose that the economic policy uncertainty reacts to economic policy changes and primarily affects economic and investment activity. On the contrary, financial uncertainty relates to the uncertain reaction of the financial markets and affects financial risks. Especially credit spreads increasing with uncertainty [26]. Therefore, we define our second hypothesis as follows:

H2: the financial uncertainty effects on the financial indicators differs before and after the financial crisis

In addition, we contribute to the recent literature with constructing the new index of financial uncertainty. There are four main approaches to measure uncertainty in the recent literature. First, several indexes based on the stock market expectations measured by option evaluation, typically CBOE Volatility Index (VIX). Second, surveys monitoring disagreements among forecasters [17]. Third, macroeconometric models [13, 21, 22] or asset pricing models [11] estimating uncertainty as the latent variables. Fourth, measurements based on the textual analysis of newspapers [5, 16, 25].

We follow Baker et al. [5] and contribute to the recent literature with the financial uncertainty index employing textual analysis. We argue that economic policy uncertainty based on the news presented in newspapers endogenously reacts to shocks that cause business cycles. However, the information about the uncertainty presented in newspapers affects financial markets and changes financial uncertainty.

[1] Mendel University in Brno, Department of Finance, Zemedelska 1, 613 00 Brno, kapounek@mendelu.cz.
[2] Charles University, Institute of Economic Studies, Opletalova 26, 110 00 Praha 1, roman.horvath@gmail.com.

The paper is organized as follows. Section 2 introduces the data and employed methods. In Section 3 we provide main results and section 4 concludes.

2 Data and Methods

First, we construct new financial uncertainty index. We follow Baker et al. [5] and count number of articles containing specific keywords combination in the following selected US newspapers: Dallas Morning News, Chicago Tribune, Los Angeles Times, Miami Herald and Houston Chronicle, New York Times, San Francisco Chronicle, The Boston Globe, The Washington Post, USA Today, and Wall Street Journal. The selection of the newspapers is based on the potential to affect decision making processes at the financial markets. We use the words "uncertain" or "uncertainty" in combination with the words "financial" or "finance".

We employ ProQuest database and generate indexes based on the automated textual searches for the specific single month. That way we generate monthly indexes from January 1985 to December 2017. Thus, we cover period of great moderation in 90s and the global financial crisis beginning in 2007, including the consequences of the both periods. Following Baker et al. (2016), the index is calculated as the frequency of searched articles in each newspaper with respect to total number of articles in the selected month. The index is at monthly frequency with a mean of 100.

Second, we apply Cholesky decomposition and estimate a VAR model of the US economy, including S&P500 index (log), federal funds rate, employment (log) and industrial production index (log). The selection of the variables is based on the regressions provided by Baker et al. [5]. Following Baker et al. [5], we include three lags of all variables. We model on-standard-deviation financial uncertainty shocks and report impulse response functions. In addition, the employment was seasonally adjusted using X-13ARIMA-SEATS method developed by the Census Bureau.

Third, we provide responses of financial uncertainty shocks to the financial variables: (1) several measures of spreads (10y Treasury–fed funds spread, 10y–2y Treasury spread, TED spread, AAA corporate bonds–10y Treasury spread), (2) measures of lending activity (consumer loans and commercial and industrial loans to assets), (3) bank margins (difference between the bank loan rate and the fed funds rate), and (4) housing prices (S&P Case-Shiller Home Price Index).

3 Results

We estimate the VAR models on two different samples: January 1985–June 2007 and July 2007–December 2017 (Figure 1). The sample period of January 1985–June 2007 is characterized by the lower level of uncertainty, while the sample period of July 2007–December 2017 is characterized by more turbulent financial markets, extensive government support to US financial sector, bank failures and large-scale unconventional monetary policy measures. In addition, the sample period of July 2007–December 2017 includes a crisis period of (mid)2007-2009.

Our results show that one standard deviation shock to the financial uncertainty decreases economic activity (represented by industrial production index and employment) and stock market index. However, this response is significantly higher after the global financial crisis. The response to financial uncertainty shocks on the economic activity and stock markets before the crisis is lower. The result is in line with our expectations that the economic policy changes after the global financial crisis boosted the effects of the financial uncertainty. In comparison to responses of the economic activity and stock markets, the reaction of Fed funds rate is limited because of zero lower bound after the financial crisis.

Generally, the effects of the financial uncertainty becomes more significant 6 months after the shock and disappears after 2 years and later. The effects of financial uncertainty are longer after the financial crisis. However, the effects of financial uncertainty on the stock markets are not significant in longer period.

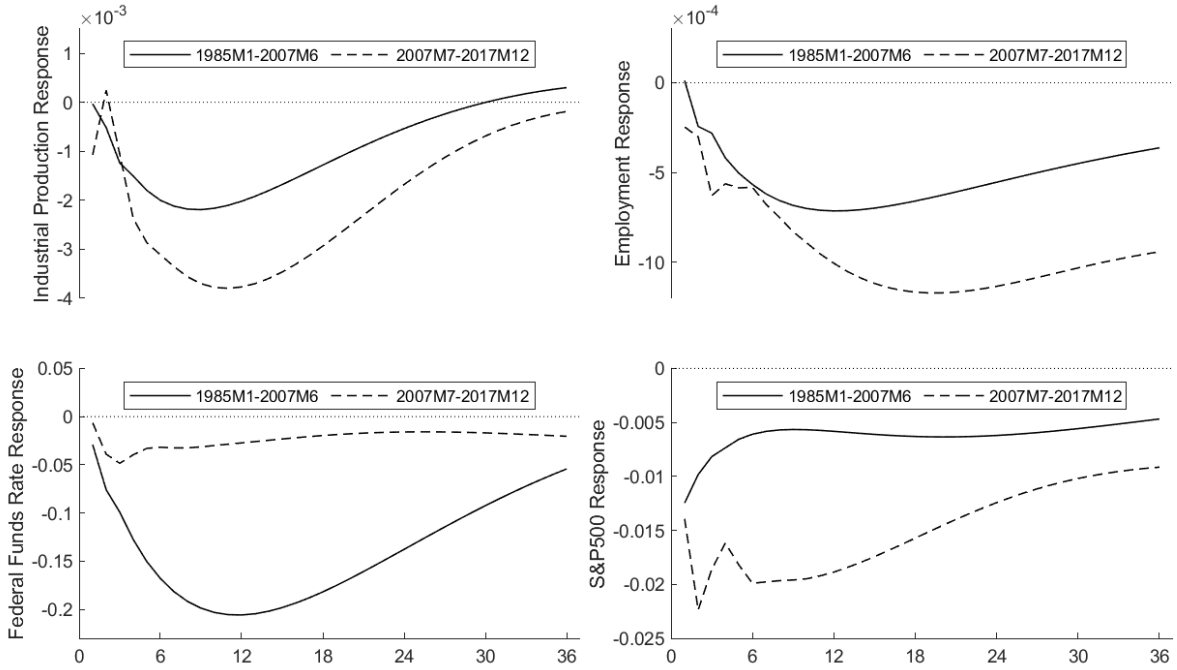


Figure 1 Responses to a Financial Uncertainty Shock before and after the Global Financial Crisis

In addition, we provide response of financial variables to financial uncertainty shock (Figure 2). We cover three bond yield spreads (10y Treasury–fed funds spread, 10y–2y Treasury spread, AAA corporate bonds–10y Treasury spread) representing yield curves, two credit risk measurements (TED spread and bank margins), housing prices (S&P Case-Shiller Home Price Index) and lending activity of the commercial banks (consumer loans and commercial and industrial loans to assets).

Our results show lower reaction of 10y spreads to financial uncertainty shocks after the financial crisis. However, the Ted Spread representing the credit risk of the global economy significantly increased 2 months after the shock. The reaction after the financial crisis is even bigger. The similar reaction is presented in case of bank margins. Both indicators are related to the asymmetric information pronounced after the global financial crisis.

On the contrary, amount of loans and house prices decreased slowly after the financial uncertainty shocks. Even, before the crisis periods, we show moderate positive effects on house prices after 6 months.

There is also difference between the response to financial uncertainty shock before and after the year 2007 in case of spreads related to federal funds. The limited reaction after the financial crisis is probably caused by the zero lower bound. However, significant positive reaction of 10y Treasury-Fed Fund Spread and 10y-2y Treasury spread signifies flattening yield curve 6-9 months after the shock. Increasing yield spreads immediately after the shock (up to 6 months) indicate the risk economic crisis. In addition, we show similar reaction of AAA corporate bonds–10y Treasury spread to financial uncertainty shock in both analyzed periods.

In line with the financial literature, spread between long- and short-term interest rates indicates negative expectations about the economic performance and credit risk in the near future. The increasing bond yields in short term and decreasing bond yields in long term indicates that investors are moving investments away from short term into long term bonds.

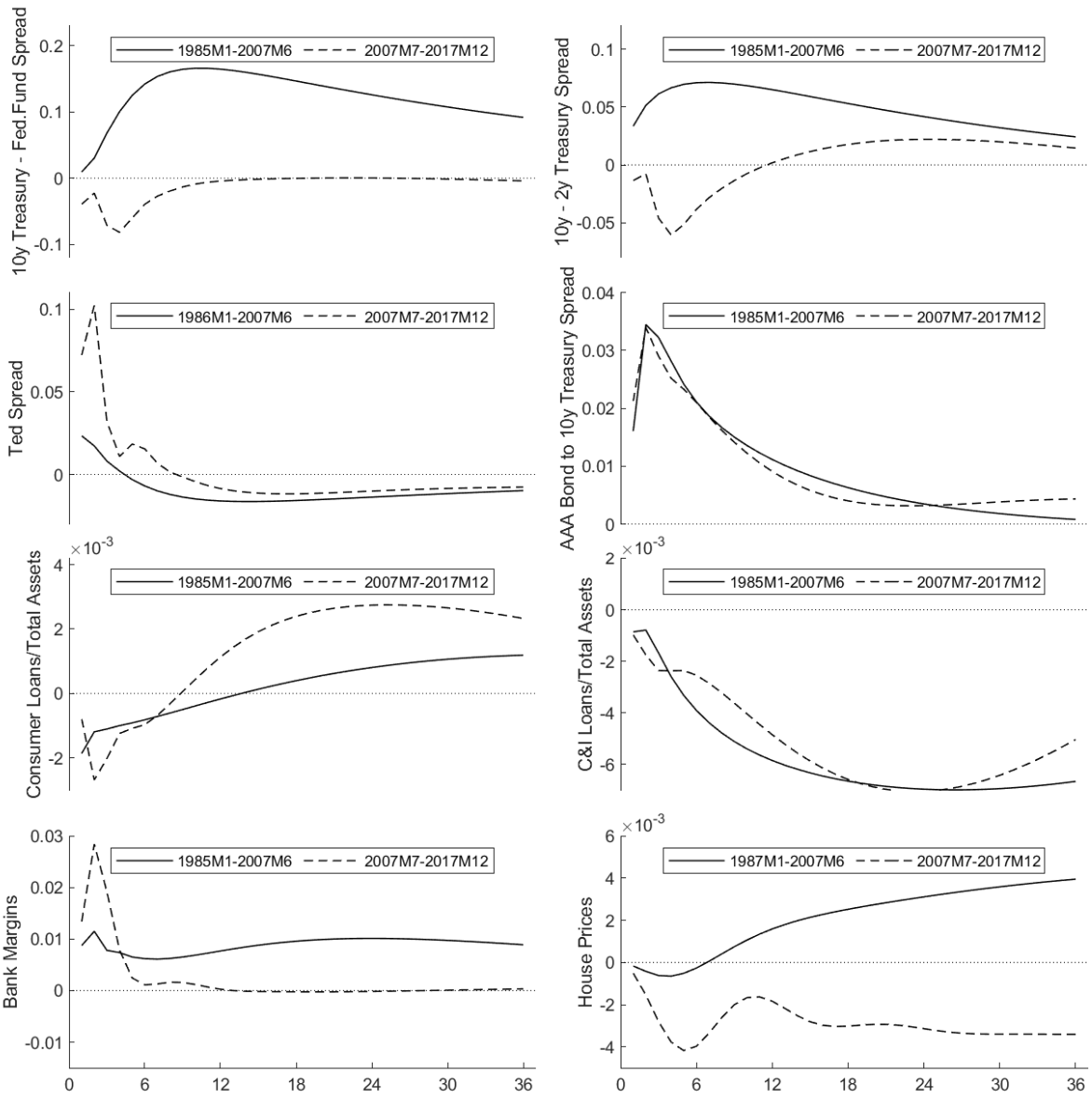


Figure 2 Responses of the Financial Sector to a Financial Uncertainty Shock before and after the Global Financial Crisis

4 Conclusions

We construct the new measurement of financial uncertainty employing textual data analysis of articles from 11 leading US newspapers. We confirm negative effects of financial uncertainty on the economic activity and stock markets. We also confirm close relation between the financial uncertainty and financial market spreads. We show increasing credit risk after the financial uncertainty shock in very short term. We also show negative effects of financial uncertainty on the bank lending activity in long period.

However, the main contribution of our paper is identification of heterogeneous effects of the selected variables to the financial uncertainty shocks before and after the global financial crisis in 2007. We confirm the hypothesis that the financial uncertainty effects on the economic activity increased after the global financial crisis. We also confirm that the financial uncertainty effects on the financial indicators differs before and after the financial crisis.

Following the theoretical literature, we argue that financial uncertainty relates to the uncertain reaction of the financial markets to news. The global financial crisis starts the period of high economic policy uncertainty represented by regulatory changes and unexpected reactions of economic policy, especially uncertainty about the effects and consequences of economic policy changes.

Our results show accelerating effects of the financial uncertainty after the global financial crisis, except the limited effects of financial uncertainty on the FED reactions caused by zero lower bound. The response of financial indicators to the financial uncertainty shocks uncover significant effects on credit risk (represented by TED spread and bank margins). With respect to employing yield curves as a leading indicator of economic crisis, we can summarize that the reaction of bond yield spreads to financial uncertainty shock is very similar to pessimistic expectations of the investors about the economic development in the near future.

Acknowledgements

We are grateful for the helpful comments we received from Iftekhar Hasan, Nicholas Bloom, Evžen Kočenda, Julius Horvath and participants of the MME conference. The authors acknowledge the financial support of the Czech Science Foundation via grant No. 22-34451S "New Methods in Pricing Government Debt: Uncertainty and Policy Implications". The authors have no conflicts of interest to declare. The usual disclaimer applies.

References

[1] Aikman, D., Kiley, M., Lee, S. J., Palumbo, M. G. & Warusawitharana, M. (2017). Mapping heat in the u.s. financial system. *Journal of Banking & Finance*, 81, 36–64.

[2] Arellano, C., Bai, Y. & Kehoe, P. (2019). Financial frictions and fluctuations in volatility. *Journal of Political Economy*, 127(5), 2049–2103.

[3] Auerbach, A. J. & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4(2), 1–27.

[4] Bachmann, R. & Sims, E. R. (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics* 59(3), 235–249.

[5] Baker, S. R., Bloom, N. & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636.

[6] Bassett, W. F., Chosak, M. B., Driscoll, J. C. & Zakrajsek, E. (2014). Changes in bank lending standards and the macroeconomy. *Journal of Monetary Economics* 62, 23–40.

[7] Basu, S. & Bundick, B. (2017). Uncertainty Shocks in a Model of Effective Demand. *Econometrica*, 85, 937–958.

[8] Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica,* 77(3), 623–685.

[9] Bernanke, B. (2018). The Real Effects of the Financial Crisis. *Brookings Papers on Economic Activity*, 49(2), 251–342.

[10] Bordo, M. D., Duca, J. V. & Koch, C. (2016). Economic policy uncertainty and the credit channel: Aggregate and bank level U.S. evidence over several decades. *Journal of Financial Stability*, 26, 90-106.

[11] Brenner, M. & Izhakian, Y. (2018). Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics,* 130(3), 503–531.

[12] Caggiano, G., Castelnuovo, E. & Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in u.s. recessions. *Journal of Monetary Economics*, 67, 78–92.

[13] Carriero, A., Clark, T. E. & Marcellino, M. (2018). Measuring uncertainty and its impact on the economy. *The Review of Economics and Statistics*, 100(5), 799–815.

[14] Gilchrist, S., Sim, J. W. & Zakrajsek, E. (2014). Uncertainty, Financial Frictions, and Investment Dynamics. *NBER Working Papers*, 20038. National Bureau of Economic Research.

[15] Hu, S. & Gong, D. (2019). Economic policy uncertainty, prudential regulation and bank lending. *Finance Research Letters*, 29, 373–378.

[16] Husted, L. F., Rogers, J. H. & Sun, B. (2019). Monetary Policy Uncertainty. *Journal of Monetary Economics*, 115(C), 20–36.

[17] Jurado, K., Ludvigson, S. C. & Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, 105(3), 1177–1216.

[18] Kapounek, S. & Kucerova Z. (2019). Historical decoupling in the EU: Evidence from time-frequency analysis. *International Review of Economics & Finance*, 60, 265–280.

[19] Kapounek, S. & Pomenkova, J. (2013). The endogeneity of optimum currency area criteria in the context of financial crisis: Evidence from the time-frequency domain analysis. *Agricultural Economics-Zemedelska ekonomika*, 59(9), 389–395.

[20] Karadima, M. & Louri, H. (2021) Economic policy uncertainty and non-performing loans: The moderating role of bank concentration. *Finance Research Letters*, 38.

[21] Ludvigson, S. C., Ma, S. & Ng, S. (2019). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *American Economic Review: Macroeconomics*, 13(4), 369–410.
[22] Mumtaz, H. (2018). *Measuring the origins of macroeconomic uncertainty.* Working Papers 864, Queen Mary University of London, School of Economics and Finance.
[23] Pastor, L. and Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3), 520–545.
[24] Pomenkova, J. & Kapounek, S. (2013). Heterogeneous distribution of money supply across the euro area. In Tvrdon, M. & Majerova I. (Eds.), *Proceedings of the 10th international scientific conference: Economic policy in the European Union member countries* (pp. 238–279). Vendryne, Czech Republic.
[25] Puttmann, L. (2018). *Patterns of Panic: Financial Crisis Language in Historical Newspapers.* Working Papers, University of Bonn.
[26] Waisman, M., Ye, P. & Zhu, Y. (2015). The effect of political uncertainty on the cost of corporate debt. *Journal of Financial Stability*, 16, 106–117.

Filtering Methods for Output Gap Estimation and the Empirical Taylor Curve: A Comparative Study

Dominik Kavřík[1]

Abstract. The empirical Taylor curve relationship between the variances of inflation and the output gap has been studied and utilized in macroeconomic policy analysis. However, the majority of existing research relies heavily on the Hodrick-Prescott filter to estimate the output gap, which may in some cases lead to biased results. This study critically examines the sensitivity of the empirical Taylor curve related to the choice of filtration technique for output gap estimation. Three alternative filtering methods are used in this analysis to highlight potential discrepancies in the resulting relationship: The Hodrick-Prescott filter, the Christiano-Fitzgerald filter and Beveridge-Nelson decomposition. These findings indicate that the choice of filtration technique significantly influences the estimated Taylor curve relationship. Consequently, this affects the evaluation of monetary policy in certain periods, calling for a more cautious approach when choosing the appropriate filtering methods for output gap estimation.

Keywords: monetary policy, output gap, Taylor curve

JEL Classification: E31, E51, C13
AMS Classification: 91B62

1 Introduction

The Taylor curve is a tool developed by John B. Taylor [10] and it is describing the negative relationship between the volatilities of the output gap and inflation. This relationship arises from the optimal monetary policy of the central bank that is assumed to be minimizing welfare loss. If the monetary policy is optimal, the observed relationship between said variance should be negative. This variability trade-off is then used as an indicator of sound monetary policy [9]. This paper approaches the issue of estimating the output gap that is required to analyze this relationship[1]. Since the output gap is defined as a deviation of the real output of the economy (real GDP) from its natural level, where all inputs are optimally allocated. Natural level and therefore the output gap is a latent variable in this analysis. It is, therefore, necessary to obtain an estimation of this unobserved variable, usually using filters or structural macroeconomic models.

Empirical researchers studying monetary policy using the Taylor curve concept, such as [7] or [9], have heavily relied on the Hodrick-Prescott (HP) filter to obtain the measure of the output gap. However, the use of the HP filter is not without controversy. [5] raised concerns about the end-point bias and spurious cycles appearing in HP-filtered series. Another concern with the HP filter is its atheoretical nature, as it is purely a statistical method that does not incorporate any macroeconomic theory or prior information about the business cycle. In contrast, band-pass filters, such as the Christiano-Fitzgerald (CF) filter [3], offer a more theoretically grounded approach to estimating the output gap. The CF filter specifically targets business cycle frequencies, based on the notion that the output gap should be associated with fluctuations at these frequencies.

2 Methodology

In this study, I aim to assess the sensitivity of the results of the previous study by Olson [9] on the choice of filtration method for the output gap. At first, the output gap will be calculated using the HP filter, which assumes deterministic time series that can be decomposed into a trend (g_t) and cycle (c_t) with a smoothing parameter λ, which is set to 1600 for quarterly data. HP filter minimizes the following objective function: $\min \left\{ \sum_{t-1}^{T} c_t^2 + \lambda \sum_{t=1}^{T} [\Delta g_t - \Delta g_{t-1}]^2 \right\}$.

The difference between the logarithm of real GDP and HP filtered trend will be the benchmark for comparison in this study.

[1] Prague University of Economics and Business, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, kavd00@vse.cz

[1] Please note that this conference paper presents a portion of the comprehensive study. The complete analysis has been submitted for publication.

Next, the HP-filtered output gap results will be compared to the Christiano-Fitzgerald (CF) filter. The CF filter is, again, used to extract the cyclical component of the real GDP, $\tilde{y}_t$, from the original GDP series, y_t as follows: $\min_{\tilde{y}_t} \sum_{t=1}^{T}(y_t - \tilde{y}_t)^2 + \lambda \sum_{t=1}^{T}(\Delta^2 \tilde{y}_t)^2$.

The final filtering approach in the comparison is the Beveridge-Nelson (BN) decomposition. Like the HP filter, the BN decomposition perceives the time series as the aggregate of a growth component (trend), denoted as g_t, and a cyclical component, c_t. Nevertheless, a significant distinction exists in the assumption about the growth component and the GDP series. Unlike other methods, the BN decomposition considers the growth component as the long-term infinite-horizon forecast of the real GDP:$g_t = \mathbb{E}_t\left[\sum_{i=0}^{\infty} y_{t+i}\right]$.

The BN decomposition also assumes that the growth component follows a random walk with drift. Consequently, the real GDP is treated as an I(1) process, resulting in Δy_t being covariance-stationary. For a more comprehensive review of this method, refer to [1]. This approach is aptly suited for difference-stationary time series. Nonetheless, it does bear a potential drawback, as articulated in [4], which is the necessity for an accurate specification of the ARIMA representation of the original time series—in this case, the real GDP. This requirement can potentially complicate the analysis.

The subsequent step in the comparative study involves obtaining the conditional variances of both inflation and the output gap estimated via the above-mentioned filters. To accomplish this, I employ the multivariate BEKK GARCH model, based on the approach in studies such as [9] and [7], to maintain consistency with the literature. The BEKK-GARCH model is estimated in the following form with the mean equation defined as AR(2) process. The order of the AR(2) process for the mean equation was chosen based on Bayesian information criterion (BIC).

$$\begin{bmatrix} \tilde{y}_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} \gamma_{\tilde{y},0} + \gamma_{\tilde{y},1}\bar{y}_{t-1} + \gamma_{\tilde{y},2}\bar{y}_{t-2} \\ \gamma_{\pi,0} + \gamma_{\pi,1}\pi_{t-1} + \gamma_{\pi,2}\pi_{t-2} \end{bmatrix} + \underbrace{\begin{bmatrix} \nu_{\tilde{y},t}\sqrt{h_{\tilde{y},t}} \\ \nu_{\pi,t}\sqrt{h_{\pi,t}} \end{bmatrix}}_{\boldsymbol{a_t}}, \tag{1}$$

where $\boldsymbol{a_t} \sim \mathcal{N}(0, \boldsymbol{H}_t)$, where $\boldsymbol{H}_t$ stands for (2×2) time-varying covariance matrix.

The estimated variances then enter the time-varying parameter (TVP) model to empirically test the time-varying Taylor Curve relationship. The TVP model defined in state-space representation as follows:

$$h_{\pi,t} = \beta_{0,\,t} + \beta_{1,\,t} h_{\tilde{y}\,t} + u_t \tag{2}$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{I}\boldsymbol{\beta}_t + \boldsymbol{\upsilon}_{t+1} \tag{3}$$

where $\boldsymbol{\upsilon}_{t+1} \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, and the evolution of $\beta_{1,t}$ is of particular interest because it measures the direction of a relationship between the output gap variance and inflation variance.

3 Data

Time series data of the real GDP and Consumer Price Index (CPI) of the United States are obtained from the database of the Federal Reserve Bank of St. Louis. Both are in quarterly frequency and are consequently used to calculate the rate of inflation and the output gap. The data set used in this study ranges from the first quarter of 1960 through the third quarter of 2021. The output gap is calculated by the previously mentioned filtering methods. For a comparison of the filtering methods and their results, see the figure below.

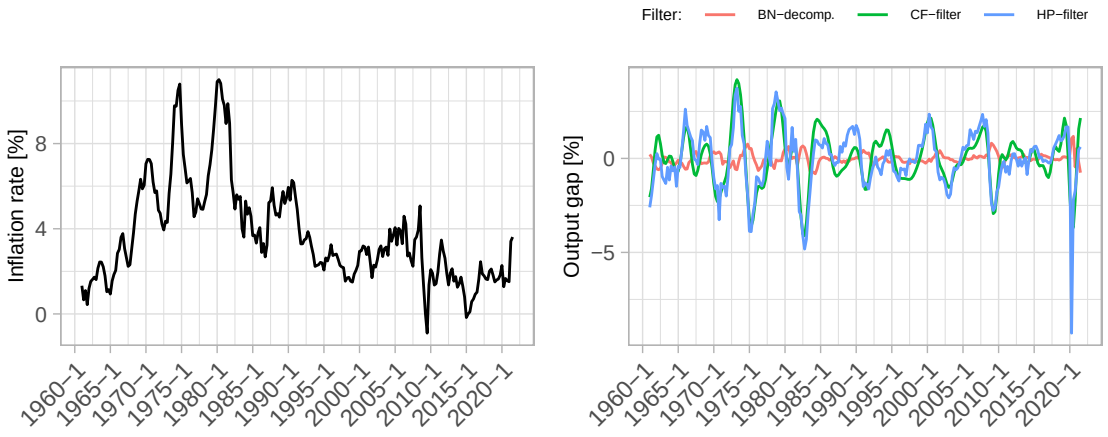


Figure 1 Time series of the output gap estimated using HP-filter (blue line), CF-filter (green line) and Beveridge-Nelson decomposition (red line).

Time series of inflation was tested for a unit root in the presence of a structural break using Zivot-Andrews [11] unit root test. The null hypothesis of a unit root was rejected on the 5% level of significance. The right plot in the figure above shows the estimated output gap series obtained using the previously mentioned filtering methods.

4 Results and Discussion

The time-varying parameter model shown in its state-space for in equations 2 and 3 is estimated using Kalman filter [6]. Time-varying parameter $\beta_{1,t}$ is of the main interest in this study because its absolute value has implications for monetary policy efficiency. If there is a trade-off relationship ($\beta_1 < 0$), it implies that the monetary policy is efficient, based on the Taylor curve theory [10], which is for this very purpose used e.g. in [9, 7, 2] and is also discussed from the historical perspective in [8]. Hence, $\beta_{1,t}$ is the key parameter for the Taylor curve estimation sensitivity assessment. The following figure shows the estimated evolution of this parameter $\beta_{1,t}$ from equation 2.

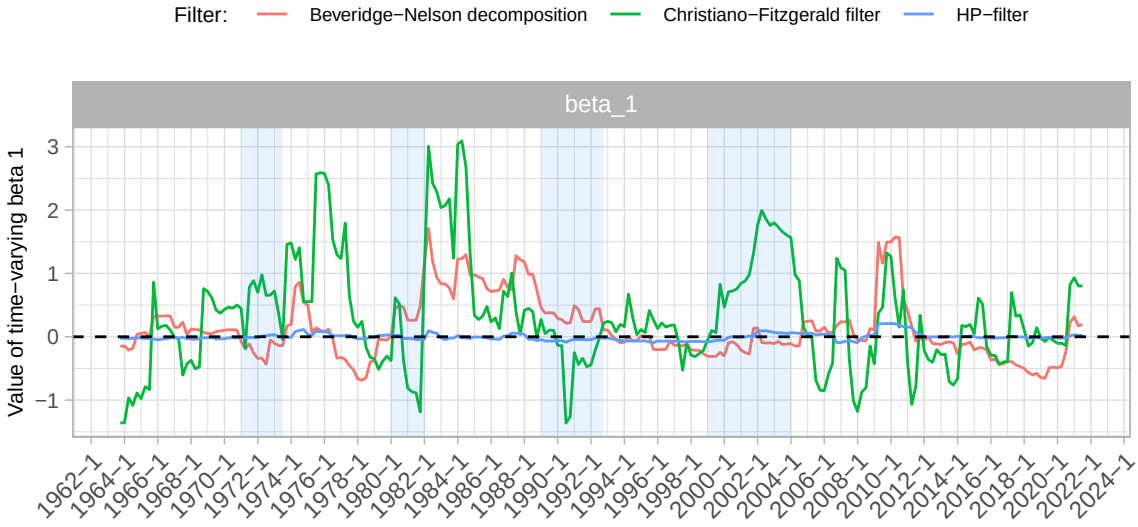


Figure 2 Time series of inflation (left) and the output gap (right) estimated using HP-filter (blue line), CF-filter (green line) and Beveridge-Nelson decomposition (red line).

Several periods show inconsistency in the estimated direction of the relationship between variances, depending on the chosen filters for the output gap estimation. Olson [9] reported a positive switch in this relationship, a finding that contradicts the results when using the Beveridge-Nelson decomposition to obtain the output gap. However,

the results derived from the CF-filtered output gap are consistent, as well as those obtained from the HP filter[2], which was the filtering used in [9].

The tenure period of Federal Reserve (Fed) Chairman Alan Greenspan (1987-2006) was characterized as "very stable" in [9], which was supported by a consistently negative relationship between the variances. However, the application of a Time-Varying Parameter (TVP) model with a CF-filtered output gap challenges this view, as this consistency is disrupted in 1992 and 1999. Furthermore, when applying the TVP model with a BN-decomposed output gap, a consistent decline in the effect size is evident from the onset of Greenspan's tenure. However, it only reaches negative values in 1993 (six years into Greenspan's tenure) casting further doubt on the perceived stability of this period.

The variations in the TVP model errors in connection with the different filtering methods are shown in the table below.

Filter	Error mean	Error sd	min	max
Beveridge-Nelson decomposition	0.0029	0.0676	-0.1869	0.3551
Christiano-Fitzgerald filter	0.0017	0.0634	-0.1131	0.3629
HP-filter	0.0032	0.0682	-0.1622	0.3229

Table 1 Model Error summary for all filters.

It's important to note that while the means of the error differ depending on the filtration method employed, the standard deviations of these errors remain relatively consistent across all filters used in the TVP model. This observation underscores the impact each filter has on the mean error while maintaining comparable levels of variability. CF-filter performs the best compared to other filters. It's also the filter that does not suffer with the end-point bias compared to HP-filter, which is also discussed in [5]. Upon limiting the sample to data prior to the 2019 GDP contraction induced by the pandemic, it's observed that the average error of the Time-Varying Parameter (TVP) model employing the HP-filter is less than that using the BN decomposition. This outcome aligns with the inherent characteristics of the HP filter mentioned above. It is crucial to remember that the standard deviations listed in the table above are computed from the error itself, offering a rudimentary understanding of the overall error distribution. These should not be conflated with the diagonal elements from the covariance matrix of the TVP model.

When comparing various filtering methods for the output gap, attention should not be limited solely to the error of the TVP model. Although the average errors might exhibit similar distributions, the derived estimates of the time-varying parameter of the volatility trade-off relationship may yield inconsistent results. Thus, a comprehensive comparison necessitates more than just assessing the TVP error.

5 Conclusion

The aim of this research was to investigate the influence of the choice of output gap filtration technique on the estimation of the Taylor Curve within a time-varying parameter model framework. The results derived from three different filtration methods: the Hodrick-Prescott statistical filter, the Christiano-Fitzgerald band-pass filter, and the Beveridge-Nelson decomposition were compared in terms of the effect size and direction in the time-varying parameter model. While the mean and standard deviation of the time-varying parameter model error were comparable across all three filters, there were distinct differences in the effect size describing the relationship between the variances of the output gap and inflation during certain time periods, which implies that the filter choice can affect the resulting evaluation of the monetary policy via the Taylor Curve estimation, highlighting the need for a careful selection of the appropriate filtration technique and possibly to avoid relying on a single filtering method.

Acknowledgements

This research was supported by the Internal Grant Agency of Prague University of Economics and Business under Project F4/38/2022.

[2] This consistency might not be immediately apparent due to the scale in the above plot.

References

[1] Beveridge, S. & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics*, 7(2):151–174.

[2] Chatterjee, S. (2002). The Taylor Curve and the Unemployment-Inflation Tradeoff. *Federal Reserve Bank of Philadelphia Business Review*, pages 26–33.

[3] Christiano, L. J. & Fitzgerald, T. J. (2003). The band pass filter. *International economic review*, 44(2):435–465.

[4] Cochrane, J. H. (1988). How big is the random walk in gnp? *Journal of political economy*, 96(5):893–920.

[5] Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, 100(5):831–843.

[6] Kalman R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45.

[7] Lee, J. (1999). The inflation and output variability tradeoff: evidence from a GARCH model. *Economics Letters*, 62(1):63–67.

[8] Olson, E. & Enders, W. (2012). A Historical Analysis of the Taylor Curve. *Journal of Money, Credit, and Banking*, 44(7):1285–1299.

[9] Olson, E., Enders, W. &, Wohar M. E. (2012). An Empirical Investigation of the Taylor Curve. *Journal of Macroeconomics*, 34(2):380–390.

[10] Taylor, J. B. (1979). Estimation and control of a macroeconomic model with rational expectations. *Econometrica: Journal of the econometric society*, pages 1267–1286.

[11] Zivot, E. & Andrews, D. W. K. (2002). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, 20(1):25–44.

Sportka - "Better" strategies based on analyses of specific number combinations?

Jana Klicnarová[1], Kateřina Walterová[2]

Abstract. Lotteries are still very popular in our country despite (or perhaps because of) the unfavourable economic situation. That is why many people are still trying to find a way how to get better results in lotteries. In this paper, we will focus on the Sportka lottery. More precisely, we will discuss a model proposed by Tautermann. His model is based on statistics of different types of combinations and uses long-term historical data, and the author believes that his model ensures better results than a random bet.

Keywords: Lottery, theoretical distribution

JEL Classification: C44
AMS Classification: 91-11

1 Introduction

The problem of lotteries is widely discussed in many papers from different points of view – as a probabilistic, economic, sociological, or psychological problem. It is well-known that all lotteries, from the principal, are – in terms of Game Theory – unfair games; on the other hand, they are very popular. Many authors studied this paradox (again from different points of view) – why people are interested in a "game" even if they know very well that their expected profit is negative; for more details, see, for example, [5, 9, 4], and references herein.

This paper focuses on a mathematical approach to the lottery problem. More precisely, we focus on one specific lottery, the Czech lottery – Sportka, the Czech phenomenon in betting. We will discuss one of the offered strategies in this lottery. Sportka is the most famous Czech lottery, the oldest Czech lottery, which has run since 1957. It is unbelievable how much money people can throw around every week in this game, hoping for their success. Since many people bet on this lottery, many of them are interested in this phenomenon, studying the game's history, thinking about "winning" strategies and discussing their approaches to this problem. Many people believe a scenario exists that ensures that the bettor following this strategy has a better position among the other bettors. This paper analyzes the historical data of this game and discusses possible strategies. We also focus on one of the assumed strategies and discuss its potential benefits. Indeed, we are not alone in who decided to study this game; many papers have already been written on this topic, and also many students chose this topic for their theses. We can mention, for example, Novák, see [6], who discussed probabilities of some rare cases in Sportka and also showed why one of the first drawing devices was probably changed after only four years of use.

1.1 Sportka Rules

At the beginning of our paper, let us briefly recall the rules of the Sportka game; see the official web [8]. However, most Czech inhabitants know them quite well since schools often use the game as an example of probabilities. The rules of Sportka are pretty simple. The only thing that the bettor has to do is guess a combination of six out of forty-nine numbers that will be drawn in the next game round. The game round takes place every Wednesday, Friday, and Sunday under the supervision of the notary. Each round has two draws, Draw no. 1 and Draw no. 2. At each draw, six numbers from 49 and one additional number are drawn. (From a combinatorics point of view, it is the choice without returns, and ordering numbers is unimportant.) On the Sportka ticket, there are ten columns, and the bettor may choose one to ten columns to bet. The bettors can also choose an additional game, "Chance," ensuring the potential win is higher if Jackpot is guessed. Chance is an extra game of Sportka. In the extra game Chance, the bettor bets on the ending digits of the ticket's serial number. To win the Superjackpot instead of the Jackpot, a bettor must bet the full ticket, and the ending number of his ticket must match the last number of Chance. If the bettor guesses all six numbers correctly but the ending serial number does not match, or he did not bet the full ticket, he wins only the Jackpot. Each bet in a column costs 20 CZK, and the Chance is for 20 CZK, too. So, the full ticket cost, with the Chance, is 220 CZK (10 columns and the Chance). It is possible to bet in selling places that are mainly in all newspaper stands, gas stations, or post offices. Also, the bettors can bet online on the official

[1] University of South Bohemia in České Budějovice, Faculty of Economics, Studentská 13, 370 05 České Budějovice, klicnarova@ef.jcu.cz
[2] University of South Bohemia in České Budějovice, Faculty of Economics, Studentská 13, 370 05 České Budějovice, waltek00@ef.jcu.cz

site of Sazka a.s. In every round, two draws are made in which six numbers are drawn and one additional number. Then there is also an extra game Chance – the bet on the six last digits of the ticket; in the combinatorics sense, it is a choice with repetition, and the ordering is important. In the following text, we analyze only wins in the game Sportka. Some important information is summarized in the following table 1.

Winning order	No. of guessed numbers	Share of principal	Probability	Average win prize
Superjackpot	6 + 1 from 10	0.1	1 : 69 919 080	136 769 789 CZK
1st order (Jackpot)	6	0.22	1 : 6 991 908	15 382 198 CZK
2nd order	5 + additional number	0.07	1 : 1165318	815 723 CZK
3rd order	5	0.09	1 : 28 422	24 971 CZK
4th order	4	0.12	1 : 516	619 CZK
5th order	3	0.4	1 : 28	113 CZK

Table 1 Theoretical and historical values

Based on the table above, six different winning orders can be won in this game. To win the lowest prize, winning the 5th order, the bettor has to guess three numbers that were drawn. On the other hand, to win the Superjackpot, the bettor has to bet a full ticket, correctly guess all of the six drawn numbers, and have the same ending number of the ticket as the last number of the supplementary game Chance was drawn. As we already mentioned, Sportka is an unfair game in the sense of game theory since only half of the bettors' deposits are used as a game principal; i.e., only half of the deposits are used for paying out the winners. It means the expected loss is $1/2$ of the bet amount. Every winning order has given a part of the game deposit that goes to the winners of that winning order; see the table 1. For example, 12 percent of the game deposit is divided among the winners of the 4th order. If more winners are in the given winning order, the amount is equally divided among them. So, when a bettor wants to win as much as possible, he needs to guess the numbers correctly and the combination that no one else has picked. The average win prizes are also shown in the table above 1.

1.2 Winnings in Sportka

According to the rules of Sportka, the bettor wins if he guesses at least three out of six drawn numbers. The value of his profit is not fixed; it depends on the game principal. Therefore, it is clear that, in terms of game theory, Sportka is not a fair game. As we can see from the table 1, forty percent of the game deposit is distributed among the bettors who win the 5th order win, i.e., who guess three numbers out of six drawn. The probability of such a win is 3.57 percent. On the other hand, a similar amount of money (more than 32 percent) is distributed among the winners of the Jackpot and Superjackot, where the probability of winning is $7.15 \cdot 10^{-8}$.

The first draw of the lottery Sportka took place on April 22nd, 1957. To this day (May 2023), almost ten thousand combinations have been drawn and could have made a millionaire out of someone. Out of all the possibilities, only one combination of six numbers was replicated during the whole history of Sportka. "The lucky combination" is 5, 14, 27, 31, 46, and 47 [8]. It first occurred on February 17th, 1960. And the second time, the draw took place on March 8th, 2009. If we look more carefully at the possibility of repetition, we can compute the probability that some of the already drawn sextuples will be replicated in the next draw. As was already mentioned, from the beginning of the Sportka to nowadays, approximately 10 000 draws have been done. The probability that one of the already drawn sextuples will appear in the next draw is $10\,000/\binom{49}{6}$, i.e., $7.15 \cdot 10^{-4}$. For example, the probability that some of the drawn sextuples will be repeated in the following three years (all drawn numbers, both draws are included, three rounds a week are supposed) is one-half. Let us discuss the probability that one of the already drawn numbers will be repeated (now, one from 10 000 numbers and the number is still growing up (after three years, it will be close to 11 000 sextuples)); however, it is not possible to bet so many numbers for a player. From this point of view, all hope of a bettor of reaching a jackpot during his life seems odd, even if we admit to betting more tickets on every possible date.

2 Theoretical distribution of results

First, it is necessary to mention that in the following theoretical part, the Sportka is supposed to be a fair game from the probability point of view, i.e., in the sense that all balls (numbers) are the same, the probability of drawing anyone is the same for all of them. Later, we will discuss this assumption and discuss the verification of it. It is clear that the lottery operator aims to have such a system that the lottery is fair from the probability point of view, i.e., no difference exists among the possible strategies. Hence, if we suppose such an ideal case, then it is

well-known that the probability of guessing six drawn numbers in the given draw is $1/\binom{49}{6}$, i.e., the probability of guessing all six numbers at least at one of two draws is $2/\binom{49}{6}$, which is $1.43 \cdot 10^{-7}$. Hence, from the condition for Superjackpot, we easily derive that the probability of Superjackpot is approximately $1.43 \cdot 10^{-8}$. It is also well-known that the probability of guessing k numbers $0 \leq k \leq 6$ follows hypergeometric distribution, i.e.

$$\mathrm{P}(X = k) = \frac{\binom{6}{k} \cdot \binom{43}{6-k}}{\binom{49}{6}},$$

where X goes for a number of guessed numbers.

3 Strategy based on historical data

As mentioned above, the operator aims to have a fair game from a probability point of view; on the other hand, many people believe that Sportka is not a fair game in this sense. Therefore, they think that some "optimal" strategy exists. One of the most trusted "optimal" strategies is based on historical data. On the other hand, best to our knowledge, many researchers (for example [7]) and students in their diploma thesis have tried to find some systematic element in the drawn numbers, but, best to our knowledge, no such influence has been proved, no systematic impact has not been approved. We also decided to test the hypothesis of the supposed random distribution of drawn numbers (the hypothesis was that the frequencies of drawn numbers are 6/49 and the alternative, that the distribution of drawn numbers differs. To run such a test, we applied χ^2 test of goodness of fit on data from 2004 to 2022, and we, unsurprisingly, also could not reject the hypothesis of the supposed theoretical distribution of drawn numbers. The p-value is too high for both draws and very similar for individual draws. Therefore, there is no reason to think that Sportka is not a fair game in terms of possible numbers prediction. The following graph also shows the frequencies of drawn numbers. We can also remark that if we decide to test the frequency of each number separately, the results fully meet the expectations; for the significant level of 0.01 must be frequencies outside of the interval $[250, 326]$, for 0.05 outside of $[253, 313]$.

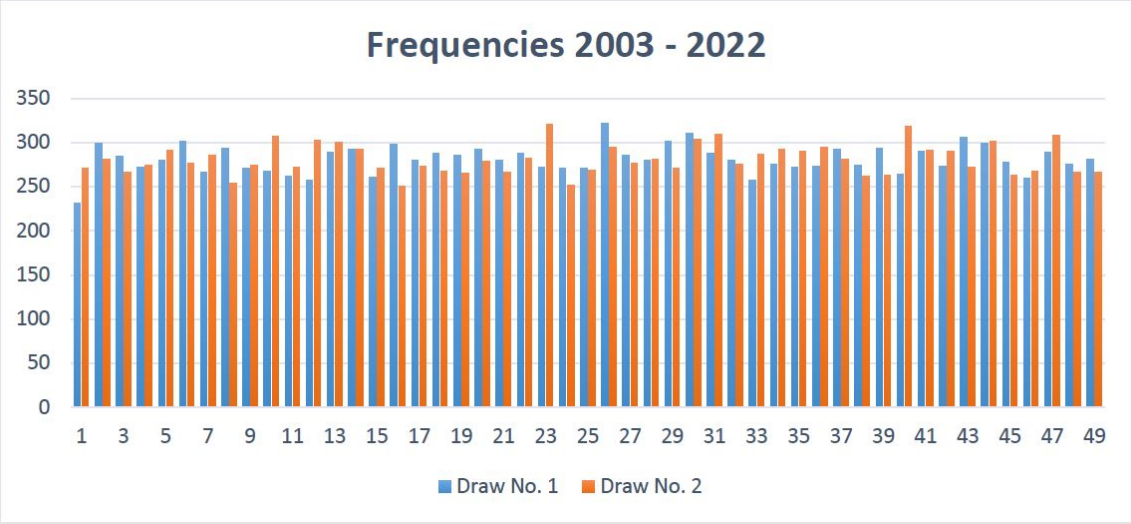


Figure 1 Frequencies of numbers

There are two main branches of theories based on historical data. The part of bettors prefer the numbers with the most frequent occurrences in history; they believe that the balls are not the same and that the higher occurrence rate in history ensures a higher probability of being drawn in the future. On the other hand, the other part of bettors prefers to bet to the numbers with the lowest occurrence rate. Their idea says that the long-run probabilities must be the same for all numbers; hence, they derive that so-far lower occurrence rates predict a higher likelihood of occurrence in the future. Let us recall our remark about the probability of repetition.

If we look at the probability of possibility to achieve the jackpot if we keep one number combination, then after 13 991 draws, the probability will rise above one per thousand.

Despite this, many people still believe they can improve their betting strategy. One of the presented strategies is a strategy developed by Tautermann and presented on his website www.loterino.eu and in his journal Loterino. In this paper, we focus on Loterino's strategy and will discuss its idea and results.

4 Evaluation of consider strategy

Tautermann decided to study the history of drawing numbers in detail. More precisely, he decided to follow specific combinations of numbers and particular characteristics of drawn sextuplet and derive conclusions about future occurrences of such combinations and characteristics. His strategy is based on the combination of the above-mentioned strategies. He follows occurrence rates of specific combinations and characteristics and prefers the combinations and characteristics with the most often occurrences; on the other hand, he prefers such combinations which have not appeared for a longer time. Based on his experience, Tautermann believes that his strategy ensures higher returns on average than a random choice of numbers. His results and recommendations Tautermann summarize in Loterino journal, see [1], where he presents which combinations have which probabilities of future appearance – he gives weights of the suitability of choice of the combinations. He updates his recommendation usually two a month. Therefore, it is necessary to mention that his recommendation is not in the form of which sextuple should be chosen, but in the form of type as choose the higher number 30, use three consecutive numbers at the last positions, use two big and four small numbers and so on. Therefore, there is no definite recommendation on which combination to choose, only the group of recommendations on what conditions the chosen numbers should meet.

To compare strategies, it is necessary to define a measure of success. In this lottery, there are two main ways how to measure success. One is to measure the rate of wins of the first order, second order, and so on. The second one is to measure the amount of financial wins. Unfortunately, these two measures are not equivalent since the wins of given orders are not fixed. They depend on a game principal, as was mentioned above. Therefore, the prize depends on the number of bet columns in a given game round and the number of successful bettors in the given level. With higher prizes – jackpots – it is more complicated because if not all money is divided among the winners in the round (nobody gets some winning order), the rest is removed to the fond for jackpots for the next rounds. Therefore, it is clear that the prize depends on the number of bettors (more precisely on the number of bet columns), and jackpots depend on results in the previous games. (In fact, it is much more complicated because there is a high level of possible jackpot funds. Therefore, if the jackpot fund is higher than the given level, some specific new rules get into the game.) Let us remark that the distribution of prizes is done from the deposit according to the key given in table 1. Therefore, for the prize at the specific level, it is important how many people guessed the given number of numbers. As more people guess as smaller profit they will have, they will divide the fixed prized among more people.

Hence, it is clear that there is no way to evaluate possible profit, and in the following, we will compare strategies according to changes to receive a prize of a given level. To do it, we will simulate the game of a "normal" player (he chooses his numbers randomly (we suppose that the expected profit from really random choice and choice based on preferable numbers is the same) and a "Loterino" player, a player who follows Loterino suggestions. Loterino recommends choosing one (or combining more) of suggested strategies and then keeping the preferred strategy to the first success. Loterino updates its recommendations once a week. For this paper, we select one of his predictions, and based on the chosen prediction, we elect 14 sextuples which should, according to the Loterino affirmation, ensure better results than a general choice. To compare this choice with a general one, we also elect 22 ordinary sextuples (some of them randomly (using a generator of random numbers in Excel), and for some of them, we ask somebody for "his" sextuples. We check the first success for each choice to compare the success rate of Loterino and general choices. It is well-known that the distribution of the first occurrence follows the geometric distribution with the parameter of the probability of such success. Since the possibilities of some winning orders are small, we focus on the 5th winning order and mention the 4th one.

For the following analysis, we use Loterino prediction from September 5th, 2022, see [2] and for comparison, some general choices. Based on Loterino's prediction, we chose the following sextuples. We apply the recommendations with the highest weights, i.e., such that their appearance is supposed with the highest probability. Based on the analysis of sequences, we choose the recommendations on $N113100$, which says to put at the position $3, 4, 5$ three consecutive numbers and chose (9, 14, 17, 18, 19, 30). Based on so-called reverse pairs, we chose (12, 21, 23, 32, 24, 42), on upper bound (30, 26, 10, 6, 22, 24, 30), small sequences (6, 23, 24, 25, 26, 27), lower bound (9, 26, 24, 22, 14, 16), pairs (34, 35, 9, 26, 6, 16), similarly other sextuples (6, 30, 26, 34, 47, 38), (17, 18, 19, 26, 6, 24), (26, 24, 16, 6, 39, 7), (2, 3, 4, 5, 26, 30), (28, 6, 16, 10, 22, 26, 34), (13, 30, 32, 35, 37, 38), (16, 6, 27, 18, 45, 11), and (36, 40, 1, 3, 46, 28). Tautermann approved this choice as correct. As a comparison sample we chosen following sets of numbers, we asked some bettors to get some sextuples, and also used MS generator of pseudorandom numbers to get the rest: (7, 21, 24, 36, 41, 42), (3, 5, 18, 28, 47, 35), (32, 42, 21, 17, 31, 30), (14, 9, 2, 22, 6, 16), (5, 12, 24, 17, 40, 39), (2, 7, 25, 12, 23, 6), (28, 9, 23, 12, 16, 3), (47, 27, 14, 13, 1, 29), (35, 18, 26, 36, 27, 23), (10, 20, 30, 40, 15, 25), (1, 2, 3, 4, 5, 6), (8, 20, 28, 43, 40, 3), (29, 45, 18, 5, 1, 13), (6, 28, 12, 27, 29, 30),

(24, 44, 15, 46, 4, 19), (8, 17, 1, 31, 16, 36), (6, 9, 14, 27, 32, 41), (4, 14, 16, 29, 33, 45), (1, 5, 16, 13, 21, 12), (49, 48, 47, 46, 45, 44), (20, 21, 22, 23, 24, 25). To compare the results of these two groups, we, firstly, follow the first realization of the 5th winning order from September 7th, 2022, to April 23rd, 2023. Hence, we get the following results, see table 2, where the number of draws until the first success of given sextuple is recorded.

Loterino, 1st draw	Loteriono, 2nd draw	Control, 1st draw	Control, 2nd draw
120	28	17	7
17	120	18	14
120	8	21	23
16	120	120	2
7	120	1	21
7	1	14	120
35	8	15	18
120	120	46	83
7	120	14	35
8	41	81	12
61	8	32	83
120	23	6	6
2	120	32	47
18	47	46	92
		31	54
		21	25
		10	32
		42	10
		32	36
		2	54
		35	9

Table 2 Waiting for the first success

In the sample, we can see several cases when there was no occurrence of success – there were 100 draws of each number. However, in the following calculation, we need to use some values in these cases, too. We decided, in such cases, to use the number 120 for the first occurrence – the reason is that in the first 100 trial, the success was not recorded, and the 19 is the last value for which the probability that the first occurrence of the success is less than or equal to 19 is less than $1/2$.

It is well-known that the distribution of the first occurrence of success is geometric with some parameters. The best – unbias and consistent – estimation of the parameter p of the geometric distribution, see [3], is:

$$\hat{p} = \frac{1}{\bar{x}}.$$

Let us estimate parameters p for different types of choices. From the theory, we know that the theoretical value is $p = 1/28 = 0.03571$. For Loterino sextuples chosen for draw no.1, we get for draw no. 1 and draw no. 2:

$$\hat{p_{l1}} = 0.02128, \text{ and } \hat{p_{l2}} = 0.01833.$$

(Even if we decide to use only numbers 100 for situations where the success was not recorded, the estimation of the parameters will be very similar, i.e., worse than theoretical values.) Hence, we can see that the results for draw no. 1 are better than for draw no. 2, as was supposed, but both estimations are worse than the expected ones; hence there is no sense in running some tests to prove the statistical significance of the positive difference. If we take a look at the comparison sample, we get the following estimations:

$$\hat{p_{c1}} = 0.03302, \text{ and } \hat{p_{c2}} = 0.02682.$$

Therefore, these estimations are better and closer to the theoretical values than the Loterino ones. Someone could raise a question if we did not omit wins of other orders. But in our trial, no winning order of 3rd and higher order has been recorded, and only in one case in the comparison sample, the winning order of 4th order has been recorded. Therefore, if we include higher winning orders, the results will change only slightly and only on behalf of the comparison sample.

It is necessary to remark that if we decide to measure the success through the value of the profit, only 40 percent of the game deposit is distributed into the 5th order winnings and more than 32 percent (in fact, it is 32 percent at least, plus non distributed prizes from the previous rounds) is distributed into the jackpot and superjackpot. Let us also remark that, in fact, the waiting for the jackpot or superjackpot is close to the waiting for unbounded profit in the well-known St. Petersburg Paradox. Theoretically, gaining the jackpot or superjackpot is possible, but the probability of it during life is really small. So, we wrote about an unfair game with the value of the game $-d/2$, where d goes for a game deposit. If we admit that it is really rare to achieve to 2nd or higher winning order, then we recognize that, in fact, the value of the game is smaller; it is $-0.7d$. From the point of view of Loterino's prediction, there is one more important remark. It is necessary to take into account also these rare phenomena. Since the probability of the occurrence of a Jackpot or Superjackpot is really small, it is very important to focus on these situations in the sense that if such a situation takes place, the player needs to be the only one who is a winner. But, if more people follow Loterino's suggestions, there is a risk that they will have similar ideas, similar, or the same, sextuples, and, in the sense of the value of the game, they reduce their possible profits.

5 Conclusion

In the paper, we discussed some probabilities of success in Sportka. Hence, we recall how small are the chances of some valuable success. Even if it is well-known as is well-known that only one-half of the deposits are put into the game principal, many people still hope for their success and pour into the lottery lots of money. We focus on the strategy presented by Loterino. In the chosen example, the strategies based on Loterino's predictions gave worse results than the randomly chosen strategy. It is necessary to remark that we chose one prediction and analyzed it. We did not do a detailed analysis of all predictions. On the other hand, we analyzed data collected from the last twenty years, and we, as many authors before, did not find any hint of some non-randomness. If this assumption is valid (and surely Sazka does everything for it), then it is clear that no prediction can help the decision-maker since the probability of any sextuple is the same. It is also important to remark that the theory of conditional probabilities helps us to understand why the theory about an early appearance of a number not drawn for a long time fails. On the other hand, we believe that the statistics done on the databases of Sazka history can help non-statisticians to understand what the statistician theory says

Acknowledgements

The paper was supported by a grant n. EF-160-GAJU-129/2022/S. The authors thank Zdeněk Tautermann for the possibility of using his database and anonymous referee for careful reading of the paper.

References

[1] ABC Quality, s.r.o. (2023): *Loterino*. Available at: https://www.loterino.eu[cited 2023-05-20]

[2] ABC Quality, s.r.o. (2022):*Loterino*, 5. 9. 2022, ABC Quality, s.r.o., České Budějovice.

[3] Anděl, J. (2011): *Základy matematické statistiky*, Matfyzpress, Prague.

[4] Beckert, J. & Lutter, M. (2013). Why the poor play the lottery: Sociological approaches to explaining class-based lottery play. *Sociology*, 47(6), 1152-1170.

[5] Nelkin, D. K. (2000). The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109(3), 373-409.

[6] Novák, P. (2006) Statistická analýza číselné hry Sportka. *Matematika v škole dnes a zajtra*, 20.

[7] Rozkovec, J. (2012). Probability distributions in loterry games. In Proceedings *The 6th International Days of Statistics and Economics*, Prague.

[8] SAZKA, a.s. (2023): *SAZKA*. Available at: https://www.sazka.cz/[cited 2023-05-25]

[9] Statman, M. (2002). Lottery players/stock traders. *Financial Analysts Journal*, 58(1), 14-21.

Facility Layout Problem with heterogeneous Material Handling System constraints

František Koblasa[1], Miroslav Vavroušek[2]

Abstract.
Manufacturing system design and production planning are among the most influential aspects of minimizing production expenses. Efficient arrangement of workshops and path planning influences the total cost of the material handling system and can reduce production lead time.
The focus of this article is to reflect the area demand of not only workshop space but also of the material handling system. That includes space which is occupied by logistic areas next to workshop entrances and space occupied by logistics aisles.
The model of this system includes the limitation of total facility space where unequal rectangular workshops have to be placed in the open field arrangement. Facility and workshops with defined access points are connected by a material handling system with an unequal size of the exits, entrances, and aisles widths. Aisled widths are based on real-world constraints, including current Czech laws and dimension demands of manipulation tools.
A constructive algorithm with dispatching rules is designed to create a facility layout. The algorithm is tested on designed problem instance. The objective is to optimize multiple criteria such as material handling costs, accessibility maintenance costs, the shape of the utilized area etc.

Keywords: Facility Layout Problem, Open Field Layout, heterogeneous aisles, material handling system.

JEL Classification: C60, C63,
AMS Classification: 90C27

1 Introduction

Facility layout problem (FLP), where workshops (departments) are placed to facility space optimizing selected criteria, is a notoriously known industrial engineering problem [17] usually mentioned as one of the oldest ones [10] as well as the most important. Its importance is not only significant from the point of operation research for its NP-hard nature [6], but also for its impact in the real world. It is estimated that 20% up to 50% of manufacturing expenses are related to material handling costs [16].

Nevertheless, there are just a few studies which are taking into account space demand for itself Material Handling System (MHS), e.g. aisles. There are in general following approaches to aisles:

- Aisles are not considered; the MHS route is defined by rectilinear [13] or contour distance [2].
- Aisles are considered, its width is inflated to workshop space [5].
- Aisles are considered, workshops have defined offset around their space [4] or its *x*,*y* sides [5].
- Aisles are taken into account, space is considered in form of one central aisle [1].
- Aisles – material handling areas are attached along the whole edge of the rectangle according to inputs and outputs. Communications (main aisles) including positions are predefined [11]
- Aisles (horizontal and vertical) are considered together with width in form of grid [14].

This paper introduces constructing heterogeneous width aisles, which is not addressed by before mentioned, by maximal free space method we used in the bin packing problem [9] and simultaneously used by Goncalves and Resende [3] as well as Paes et al. [12] for FLP to be later used for aisles construction in our research.

Paper is focusing on optimizing material handling cost of open field facility layout with unequal workshops. However, there are also considered criteria of, available regular space for future development, aisles maintenance cost and narrowness of planned road.

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

This paper is organized as follow. Second chapter focus on problem definition as well as on definition of above mentioned objective functions. Third part presents constructive algorithm as well as decision rules used in solution generation. Fourth part presents developed test instance and discussion on definition of aisles width given by Czech standard. Last part is discussing results and objective function and decision rules influence.

2 Open fields Facility Layout Problem with unequal size workshops and heterogeneous MHS constraints.

This article focus on Facility layout problem which consist of placing N unequal rectangular workshops with defined width w_i and height h_i in to unequal facility space with unequal size (W,H). Workshops can be placed freely (1)(2) in limited space (3)(4) in matter of 90° degree rotation (u_i, v_i) where (0,0) = 0°; (1,0) =90°; (0,1)=180°; (1,1) = 270°)

$$x_i' = x_i + (1-u_i)w_i + u_i h_i \quad \forall i \tag{1}$$

$$y_i' = y_i + (1-u_i)h_i + u_i w_i \quad \forall i \tag{2}$$

$$x_i' \le W_i \quad \forall i \tag{3}$$

$$y_i' \le H_i \quad \forall i \tag{4}$$

Workshops cannot overlap (5)(6)(7)

$$l_{ij} + l_{ji} + b_{ij} + b_{ji} \ge 1 \quad \forall i < j \tag{5}$$

$$x_i' \le l_{ij} x_j + W\left(1 - l_{ij}\right) \forall i, j \tag{6}$$

$$y_i' \le b_{ij} y_j + H\left(1 - b_{ij}\right) \forall i, j \tag{7}$$

where l_{ij} =1 if cell i is placed to the left of cell j($x_i \lesssim x_j$) else l_{ij} =0; b_{ij}=1 if cell i is placed to the below of cell j($y_i \lesssim y_j$) else b_{ij}=0.

Pickup and drop off points are defined for each workshop as well as for facility (Enter/Exit). Position $x^{p(d)}$,$y^{p(d)}$ of pick-up and drop-off stations are defined as (8) and (9).

$$x_i^{p(d)} = x_i + (1-u_i)(1-v_i)P(D)_i^x + P(D)_i^y u_i(1-v_i) + (w_i - P(D)_i^x)(1-u_i) + \left(h_i - P(D)_i^y\right)u_i v_i \quad \forall i \tag{8}$$

$$y_i^{p(d)} = y_i + (1-u_i)(1-v_i)P(D)_i^y + (w_i - P(D)_i^x)u_i(1-v_i) + \left(h_i - P(D)_i^y\right)(1-u_i)v_i + P(D)_i^x)u_i v_i \quad \forall i \tag{9}$$

Pick up (w_{pun}, h_{pun}) and drop off points (w_{don}, h_{don}) as well as enter (w_{ent}, h_{ent}) and exit (w_{ex}, h_{ex}) have square shape w=h to allow all directional; in 90° degree rotation except wall of itself workshop; MHS access (see more in chapter 3). Its size (lets have generalization of access point size as w_{acc},h_{acc}) is defined by maximum width w_{pk} (aisle) of all k paths connected to access point.

$$w_{acc}, h_{acc} = max\{w_{pk}\} \tag{10}$$

Aisles with variable width w_{pk} can overlap with itself as well as with access points, however they can not overlap with space of workshops.

There were considered several FLP based criteria for optimization beside classical material handling cost *MHC* (11) defined by the intensity of material flow f_{ij}, c_{ij} cost of one unit transportation and d_{ij} distance of a path between i and j logistics access points (N pick up and drop off, M enter and K exit)

$$MHC = min\left\{\sum_{i=1}^{N+M}\sum_{j=1}^{N+K} f_{ij}c_{ij}d_{ij}\right\} \tag{11}$$

Accessibility maintenance cost *(ACMC)* can be simplified as the sum of all path area spaces defined by its d_{ij} distance between i and j logistic point and width w_{pk} of k path and a sum of all logistics access points areas (12) (workshops pick-up, drop-off and facility enter and exit).

$$ACMC = \sum d_{ij} w_{pk} c_{mp} + \sum w_{pun}\, h_{pun} c_{pu} + \sum w_{don}\, h_{don} c_{do} + w_{ent} h_{ent} c_{ent} + w_{ex} h_{ex} c_{ex} \quad (12)$$

However that neglects fact that paths can (and its desirable to be) be overlapped. For that case whole logistic system is drawn in separate layout layer, thus it is possible to measure total area of MHS separately.

Number of turns (*Nturns*) during material handling is criteria which reflects ease of transportation. It decreases speed of transport as well as quality of life of logistics. As it is mostly used for transporting valuable cargo. It can be calculated as number of number of MHS intersections for general transportation [15]. In case of our research number of changes from horizontal to vertical rectangle representing aisle and vice versa is calculated as *Nturn*.

$$Nturns = \sum_{1}^{i} \sum_{1}^{j} d_j \neq d_{j+1}\,; i = 1,2,\dots n\,; j = 1,2,\dots, m-2 \quad (13)$$

$$d_j \begin{cases} 0 \;\; if \;\; x_{i+1} - x_i = 0 \\ 1 \;\; if \;\; x_{i+1} - x_i \neq 0 \end{cases} \quad (14)$$

where *i* represents number of aisles and *j* represents number of rectangles representing particular aisle and *x* represent coordinate of aisles.

Biggest square (*MaxSq*)criteria is then evaluating maximal remaining space in form of square that is available for future use. It reflects flexibility of layout in degree of not rearranging already placed departments.

3 Constructive algorithm

Constructive algorithm (see Algorithm 1) follows generally same pattern as presented in [8]. It includes sequencing **N** of workshops to be placed. Sequence is given by dispatching rule as well as by conflict set **CS**. **CS** includes workshop which has at least one connection between P/D points with already placed objects or connection with entrance or exit to facility (En/Ex), generating free spaces **F** and checking if free space is feasible for placement by testing possible rotation **r** of workshop (0°,90°,180°,270°) in defined area. After such **rpr** positions and free spaces are found all required aisles (paths) are calculated from Pick up drop off points (P/D) as well as enter and exit to facility.

Thanks to heterogeneous width of paths w_p it can happen that newly placed workshop will have not access in full width. Previously it was not possible as **CS,** thanks to homogeneous aisle width, ensured that only workshops with available access to already placed P/D/En/Ex can be placed and at the same time aisles can overlap. So orientation nor placement can't prevent access to newly placed workshop. However as w_p differs, already placed workshop can have smaller width of path, so newly placed cannot use it (see example on Figure 1 - access point of workshop 1 does not have wide enough access through access point of workshop 10).

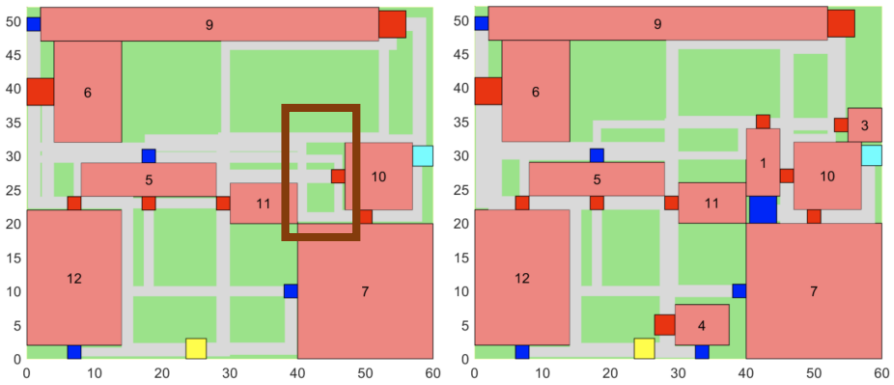


Figure 2 Access point workshop 1 small aisle width (blue)

To address this problem it was necessary to implement tabu list **TL** where solution state is placed to be banned in CS if its placement result in non-feasible solution. Backtracking **BT** mechanism is than used (steps 16-19 of Algorithm 1). Backtrack goes to state of selecting different rotation and placement to free space until none are available (see Algorithm 1 step 11). Than it goes back on the level of sequencing until it find feasible solution (see Algorithm 1 step 06).

There were defined priority rules [8] (with newly defined 1-d) to minimize search space during constructing feasible schedule. It consist of three stage decisions Sequencing-Free space selection- Placement selection (abbreviations is then used in experimental session):

1) Sequencing in step 07 is done by ordering workshops by:
   a. (IT) total decreasing intensity flow between **NB** (not planed) and **PB** (planed) logistics points.
   b. (I) the decreasing total intensity flow P/D/M/K $f^* = \sum_{i=1}^{N+M} \sum_{j=1}^{N+K} f_{ij}$
   c. (MaxA) the decreasing department area A=WH
   d. (MaxPD)maximal size of P/D of placed workshop defined by width w_{pk} of used aisle
2) Select free space by the m/M rule of the most fitting department (trying to fit largest workshop in smallest free space) in step 08 – place the department in the smallest feasible area. Option is this step is then:
   a. (F) for full search as it is defined in Algorithm 1
   b. (m/M) for most fitting
3) Selecting one placement (out of 9) and rotation in that place (0°,90°,180°,270°) by distance straight line distance d_{ij}^* instead of steps 11-13. Options are
   a. (ND) do full search in rotations and placements
   b. (AD) select the one with shortest direct distance

Pseudocode of itself algorithm for full search (x-F-ND) in placement scenario follows.

Algorithm 3 Constructive algorithm
Initialise:
create list of **PB** planned and **NB** un-planed workshops which includes **N**+**M**+**K** (access points)
add enter **M** and exit **K** points as a planned department to **PB**.
create new Bin **B** in the **Bins**
add new FreeSpaces **F** = (w; h) to **B**
Place departments
for each N placement of department **N=(w;h;f;P(x,y);D(x,y))**
  create conflict set **CS** from **NB**, which has flow intensity **f**>0 with P/D those in **PB** and are not in **TL**
  select one **N** of **NB** from **CS** to be placed in **F**
    for each F(w,h) in **B**
      for each rotation **r (w,h,a)** with angle **a**
        if F.w>=N.w and F.h>=N.h **add r** (w,h,a) to **R**
      for each rotation **r in R** excluding those in **TL**
        for each place **rp**=9
          for each r=4 rotation **in** placement **rp**
          create space assigments **rpr** =**(w;h;f;P(x,y);D(x,y))**
          find shortest path **d** of every **rpr** (P,D) and every **N** in **PB**
          if no feasible **d** is found **add** selected **rpr** state to **TL**
            if no **rpr** is available **BT** go to step 06
            elseif go to step 11
          else
            calculate MHC=f*c*d of every **P/D and M/K**
            select rpr with min**MHC** and **add** its **N in PB**
            break
    add new FreeSpaces **F** = (w; h) to **B**
    optimize FreeSpaces **in** the Bin **B** (Erase FreeSpace, which is only subspace)

4 Test instance

There was defined new instance 6-1-1 with 6 workshops to be placed in space of facility (see Figure 2). Sizes of workshops (w,h), positions of P/D(x,y) points and intensity (f_{ij}) of the transportation flow are used from [18]. Size of facility place as well position of En/Ex access points and intensity flow was defined theoretically. Width w_{pk} of aisles were set base on intensity of material flow and by ČSN 26 9010 standard [7].

Intensity f_{ij}=<1;2> and f_{ij}=<3;4> (see Figure 2) uses one direction path of w_{pk}=*0.4*+w_b where box of size w_b =0.6 m ; f_{ij}=<3;4> palette size of w_b =1.2 m) is transported by carriage. Intensity f_{ij}=<5;6> and f_{ij}=<7;8> is for two direction version w_{pk}=*0.8*+w_b of previous system. Intensity f_{ij}= 9 represent main road with one sidewalk w_{pk}=*1.2*+*2*w_b and f_{ij}= 10 with two sidewalks w_{pk}=*1.6*+*2*w_b where palette w_b=1,2 m is transported. Size of P/D (w_{pun}, h_{pun}) resp. (w_{don}, h_{don}) and En/Ex (w_{ent}, h_{ent}) resp. (w_{ex}, h_{ex}) are defined as maximal w_{pk} size used in respected access point (eg. En as well as Ex has size 4m access point).

Costs in case of tested objective functions (c_{ij}, c_{mp}, c_{pu}, c_{do}, c_{en}, c_{ex}) = 1 so objective functions *MHC* (11) and *ACMC* (12) are using dimension unites.

Sized and P/D locations [m]					P/D intensity flow *f*								P/D aisle width [m] w_{pk}							
N	*W*	*H*	P (x,y)	D(x,y)	P/D	1	2	3	4	5	6	Ex	P/D	1	2	3	4	5	6	Ex
1	10	5	0;2.5	10;2.5	1	0	1	2	1	2	3	3	1	0	1	1	1	1	2	2
2	5	5	0;2.5	5;2.5	2	5	0	1	2	1	2	5	2	1,6	0	1	1	1	1	1,6
3	20	5	10;0	10;5	3	2	3	0	3	2	1	3	3	1	2	0	2	1	1	2
4	8	6	4;0	4;0	4	4	0	0	0	1	2	4	4	1	0	0	0	1	1	1
5	12	4	0;2	6;0	5	1	2	0	5	0	0	5	5	1	1	0	1,6	0	0	2
6	9	6	4.5;0	0;3	6	0	2	0	2	10	0	10	6	0	1	0	1	4	0	4
En/Ex	40	26	0;8	17;26	En	0	3	2	5	10	3	0	En	0	2	1	1	4	2	0

Figure 2 6-1-1 instance

5 Experimental results

Algorithm 1 is tested on instance 6-1-1 (Figure 2) using dispatching rules and free space and position search described in last chapter. There are tested (Figure 3) two sets of tests evaluating four main criteria (minimization of *MHC* (11), *(ACMC)*(12), *NTurns*(13), and maximization of *MaxSq*). The first, full **rpr** position search, is scheme where all Sequencing rules are testes (*x*=IT;I;MaxA;MaxP/D) while Free space and position selection is set to *x*–(F)-(ND). Fast approach using same as before (IT;I;MaxA;MaxP/D) while Free space and position selection is set to *x*–(m/M)-(AD). Figure 4 is showing results of four objective functions given by mentioned dispatching rules. Second (ranking) part evaluates how these rules performed in each criteria to be then summed by total rank (smaller better). This substitutes in deep multicriteria evaluation which would be necessary, however it is out of scope of this paper. Rank is calculated as position in tournament where the best performance takes 1st place and the worst 8th. If there is shared position worse of them (eg. first two rules which share 1st and 2nd position in *MHC*, *ACMC* and *MaxSQ* are ranked as 2nd place).

Seq-FreeS-Poss	*MHC [m]*	*ACMC [m^2]*	*Nturns [1]*	*MaxSQ [m2]*	*MHC*	*ACMC*	*Nturns*	*MaxSQ*	Rank
IT-F-ND	6906,07	260,6	58	310	2	2	4	2	10
I-F-ND	6906,07	260,6	58	310	2	2	4	2	10
MaxA-F-ND	9096,99	382,39	54	238	5	5	2	4	16
MaxP/D-F-ND	7434,75	300,42	42	274	4	3	1	3	11
IT-m/M-AD	9731,9	434,34	230	136	7	7	8	8	30
I-m/M-AD	9731,9	434,34	230	136	7	7	8	8	30
MaxA-m/M-AD	12066,25	519,75	78	208	8	8	6	5	27
MaxP/D-m/M-AD	6975,7	328,58	64	192,4	3	4	5	6	18

Figure 3 6-1-1 instance

The best instance results were obtained with IT and I schemes with F-ND searching schemes as expected. However there are very promising results of rule MaxP/D-m/M-AD. Reason can be that MaxP/D rule reflects not only potential *ACMC* but also *MHC* as size of P/D and En/Ex are given by transport intensity. If access point size is given by other aspects, results can be different. MaxP/D was also successful in number of turns (*NTurn*) during transportation (see Figure 4). It has to be taken in account that algorithm was tested only on one instance so assumptions are limited to this case only. m/M rule which is focusing on selecting most fitting free space is the worst overall which is not surprising in all criteria but *MaxSQ*. m/M works generally well only in the combination with sequencing by MaxP/D outperforming rest of *x*-m/M-AD.

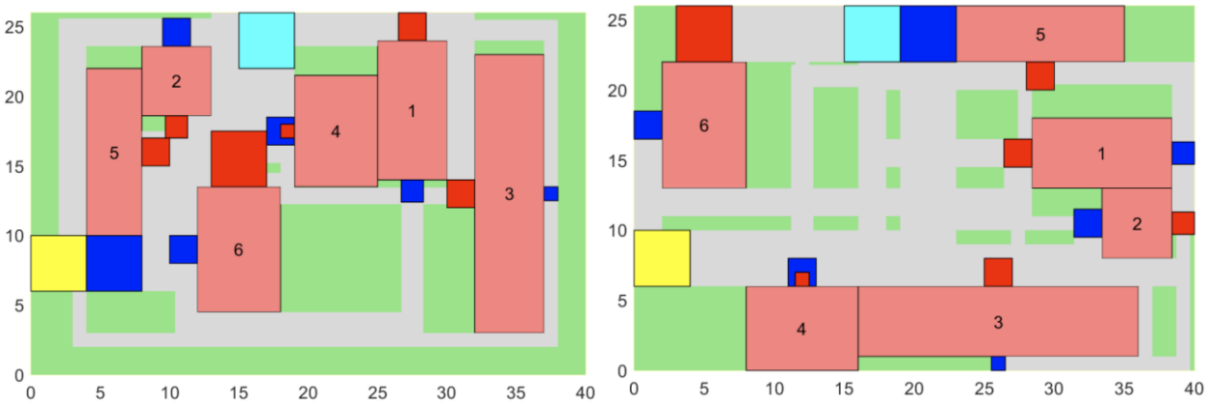


Figure 4 Layout solutions IT-F-ND and MaxP/D-F-ND

6 Conclusion

Presented research shoves important aspect of path planning in FLP as it is complication given by non-equal path widths which can cause non feasible solutions. Backtracking and tabu list was applied to solve this problem. It is shown that designed decision rules can have impact on results, however, it is limited to one test instance. Nevertheless, the goal was not to optimize FLP, but to design constructive algorithm to generate feasible solution with given constraint. Future research will focus on optimization of path planning-merging algorithm as well as on multicriteria decision making for FPL. More instances reflecting practical tasks will be developed.

Acknowledgements

This publication was written at the Technical University of Liberec, Faculty of Mechanical Engineering with the support of the Institutional Endowment for the Long Term Conceptual Development of Research Institutes, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2023. The research reported in this paper was supported by institutional support for nonspecific university research.

References

[1] Allahyari, M. Z. & Azab, A. (2018). Mathematical modeling and multi-start search simulated annealing for unequal-area facility layout problem. *Expert Systems with Applications*, *91*, 46–62.

[2] Friedrich, C., Klausnitzer, A. & Lasch, R. (2018). Integrated slicing tree approach for solving the facility layout problem with input and output locations based on contour distance. *European Journal of Operational Research*, *270*(3), 837–851. https://doi.org/10.1016/j.ejor.2018.01.001

[3] Gonçalves, J. F. & Resende, M. G. (2015). A biased random-key genetic algorithm for the unequal area facility layout problem. *European Journal of Operational Research*, *246*(1), 86–107.

[4] Guan, C., Zhang, Z., Liu, S. & Gong, J. (2019). Multi-objective particle swarm optimization for multi-workshop facility layout problem. *Journal of Manufacturing Systems*, *53*, 32–48.

[5] Hu, G. H., Chen, Y. P., Zhou, Z. D. & Fang, H. C. (2007). A genetic algorithm for the inter-cell layout and material handling system design. *The International Journal of Advanced Manufacturing Technology*, *34*(11), 1153–1163. https://doi.org/10.1007/s00170-006-0694-0

[6] Jakob, K. & Pruzan, P. M. (1983). The simple plant location problem: Survey and synthesis. *European Journal of Operational Research*, *12*(36–81), 41.

[7] ČSN 26 9010 (269010), Pub. L. No. § 3 zákona č. 142/1991 Sb. (1993).

[8] Koblasa, F. & Vavrousek, M. (2022). Facility Layout Problem with Logistic Constraints. In H. Vojackova (Ed.), *40th International Conference Mathematical Methods in Economics 2022* (pp. 180–186). Coll Polytechnics Jihlava.

[9] Koblasa, F. Vavroušek, M. & Manlig, F. (2015). Two-dimensional Bin Packing Problem in batch scheduling. *Proceedings of 33rd International Conference Mathematical Methods in Economics*, *1*, 354–359.

[10] Koopmans, T. C. & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica: Journal of the Econometric Society*, 53–76.

[11] Kubalík, J., Kurilla, L. & Kadera, P. (2023). Facility Layout Problem with Alternative Facility Variants. *Applied Sciences*, *13*(8), 5032. https://doi.org/10.3390/app13085032

[12] Paes, F. G., Pessoa, A. A. & Vidal, T. (2017). A hybrid genetic algorithm with decomposition phases for the unequal area facility layout problem. *European Journal of Operational Research*, *256*(3), 742–756.

[13] Park, H. & Seo, Y. (2019). An efficient algorithm for unequal area facilities layout planning with input and output points. *INFOR: Information Systems and Operational Research*, *57*(1), 56–74.

[14] Pourvaziri, H., Pierreval, H. & Marian, H. (2021). Integrating facility layout design and aisle structure in manufacturing systems: Formulation and exact solution. *European Journal of Operational Research*, *290*(2), 499–513. https://doi.org/10.1016/j.ejor.2020.08.012

[15] Teran-Somohano, A. & Smith, A. E. (2023). A sequential space syntax approach for healthcare facility layout design. *Computers & Industrial Engineering*, *177*, 109038. https://doi.org/10.1016/j.cie.2023.109038

[16] Tompkins, J. A., White, J. A., Bozer, Y. A. & Tanchoco, J. M. A. (2010). *Facilities planning*. John Wiley & Sons.

[17] Trebuna, P., Pekarcikova, M., Duda, R. & Svantner, T. (2023). Virtual Reality in Discrete Event Simulation for Production–Assembly Processes. *Applied Sciences*, *13*(9), 5469. https://doi.org/10.3390/app13095469

[18] Welgama, P. S. & Gibson, P. R. (1993). A construction algorithm for the machine layout problem with fixed pick-up and drop-off points. *The International Journal of Production Research*, *31*(11), 2575–2589.

Estimation Procedure for Complex Model with Spatial and Temporal Features

Martin Konopásek[1]

Abstract. The relations between macroeconomic variables distributed over time and space could be very complex. Therefore, when performing empirical analysis, it is important to take into account as many assumed features of the examined variables as possible. This article introduces dynamic panel regression model that account for numerous features that one could expect within modelled relationships and provides estimation procedure to obtain consistent estimation of corresponding parameters. This complex model allows for modeling spatial and temporal dynamics, as well as spatially heterogeneous responses while using endogenous regressors in one framework.

Keywords: linear panel model, spatial and temporal dynamics, spatial heterogeneity, random response parameters, endogenous regressors

JEL Classification: C13, C23, C26, C51, C61
AMS Classification: 60-08

1 Introduction

The article adopts the model and estimation approach of Elhorst and Zeilstra (2007) [4]. Their spatial dynamic model, tailored for cross-country regional analysis, operates on panel data with hierarchical cross-sectional dimension with two levels (e. g. countries and regions). This model captures both country-specific traits and features common to all spatial units. Specifically, it incorporates country-specific exogenous spatial effects through SAR processes, considers various variables at regional and national levels (which can be exogenous or endogenous), accounts for country-specific variances of identically adn independently distributed (IID) components, and introduces a country-specific random component for response parameters of exogenous country-level variables. The temporal dynamics, represented by AR(1) process specification, share autoregressive coefficient which is common for all spatial units.
This article provides an extension of their model by incorporating endogenous spatial effect, achieved by formulating explained variable as country-specific SAR process. Additionally, to account for spatial-temporal dynamics, explained variable lagged simultaneously in space and time with one common parameter is included. Especially former extension could be very interesting since it allows for modelling country-specific spatial responses of explanatory variables. The article is then simply structured as follows: the second section formally describes the model and provides the corresponding assumptions, while the third section is devoted to model estimation.

Notations: subscript i belongs to cross-sectional units and N is their total number; subscript c belongs to countries and C is their total number; subscript t belongs to time observations and T is their total number; bold subscripts denote names of vectors and matrices, for square matrices they also describe their dimension; $\boldsymbol{I}$ is identity matrix and $\boldsymbol{E}$ is square matrix of ones.

2 Model Description

As mentioned above, introduced model is designed for hierarchical two-level cross-sectional dimension, where first level is represented by countries and second level by regions. Further, time dimension is assumed to be short compared to the cross-sectional dimension, which could be formally expressed by assumption:

Assumption 1. *N goes to infinity, while T is fixed.*

One of the main features of described model is, that it incorporates endogenous and exogenous spatial effects, which are furthermore assumed to be heterogeneous across countries. Specifically, for every country $c \in \{1, ..., C\}$, with certain number of spatial units (regions) denoted by N_c, it is assumed specific spatial weight matrix $\boldsymbol{W}_c$ defined by following assumptions:

[1] University of Economics in Prague, Department of Econometrics and Operations Research, Winston Churchill Square 1938/4, 130 67 Praha 3 – Žižkov, konm04@vse.cz

Assumption 2. *For every country $c \in [1, ..., C]$, spatial weight matrix $\boldsymbol{W}_c$ is gained by row-normalizing connectivity matrix $\boldsymbol{C}_c$ (such as $w_{cij} = c_{cij}/\sum_j c_{cij}$), while for elements of matrix $\boldsymbol{C}_c$ applies:*

$$c_{cij}(d_{cij}) = \begin{cases} 0, & \text{if } i = j \\ 0, & \text{if distance criterion } d_{cij} \text{ is not met} \\ 1, & \text{if distance criterion } d_{cij} \text{ is met,} \end{cases} \tag{1}$$

where d_{cij} is some pre-formulated function of geographical distances between spatial units.

Assumption 3. *For rows and columns of the connectivity matrix $\boldsymbol{C}_c$ applies:*

$$\sum_{i \in [1,..,N_c]} |c_{cij}| \leq B < \infty \; \forall \; j \in [1,..,N_c] \text{ and } \sum_{j \in [1,..,N_c]} |c_{cij}| \leq B < \infty \; \forall \; i \in [1,..,N_c]. \tag{2}$$

In this set up, only interactions between regions within one country are accounted for thus weight matrix for complete model ($\boldsymbol{W_N}$) has block diagonal structure with C blocks and c-th block given by $\boldsymbol{W}_c$. This spatial structure allows for easier model construction (and parameter estimation) and could be justified by the assumption, that regions within one country are more likely to interact with each other than those which belong to different countries due to absence of social, cultural and language barriers as well as institutional differences. Although assumption of zero spatial dependence between neighbouring regions of different countries may be unrealistic in some cases, these kind of interactions are at least assumed to be weaker, thus described spatial weight structure could be consider as good approximation. Structure of matrix $\boldsymbol{W_N}$ allows to model examined variable as Spatial Autoregressive (SAR) process with C country-specific spatial parameters $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_C]^\intercal$. Specifically, dependent variable could be rewritten as:

$$\boldsymbol{y}_{ct} = \lambda_c \boldsymbol{W}_c \boldsymbol{y}_{ct} + X_{ct}, \tag{3}$$

where X_{ct} generally represents every other observed and unobserved components assumed within the process. From this specification it could be seen that for every country there must exist inverse of the matrix $\boldsymbol{A}_c = (\boldsymbol{I_{N_c}} - \lambda_c \boldsymbol{W}_c)$ for stability of the model. Additionally, SAR process within unobserved component (exogenous spatial effects) could be modelled in similar way, again with country-specific spatial parameters $\boldsymbol{\rho} = [\rho_1, ..., \rho_C]^\intercal$ and corresponding matrices $\boldsymbol{B}_1$,...,$\boldsymbol{B}_C$, where $\boldsymbol{B}_c = (\boldsymbol{I_{N_c}} - \rho_c \boldsymbol{W}_c)$ for which invertibility conditions are applied as well. In general, matrices $\boldsymbol{A}_c$ and $\boldsymbol{B}_c$ are defined by assumption:

Assumption 4. *For every country $c \in [1, ..., C]$, matrices $\boldsymbol{A}_c$ and $\boldsymbol{B}_c$ are invertible for any $\lambda_c, \rho_c \in \Lambda_c$, where parameter space Λ_c is compact and weight matrix $\boldsymbol{W}_c$ is defined by assumptions 2 and 3.*

To deal with temporal dynamics, common autoregressive coefficient (τ) associated with the time lag of dependent variable is included in model (since estimation procedure would with country-specific autoregressive coefficients loose some desirable properties described in third section). For dynamic completeness, also spatial-temporal component could be implemented, but as in case of autoregressive coefficient, only one common parameter (δ) is assumed, thus dependent variable with all aforementioned features would take form:

$$\boldsymbol{y}_{ct} = \lambda_c \boldsymbol{W}_c \boldsymbol{y}_{ct} + (\tau \boldsymbol{I_{N_c}} + \delta \boldsymbol{W}_c) \boldsymbol{y}_{ct-1} + \boldsymbol{B}_c^{-1} \boldsymbol{v}_{ct} + X_{ct}. \tag{4}$$

By taking $\lambda_c \boldsymbol{W}_c \boldsymbol{y}_{ct}$ to the left side of the equation 4 and solving for y, we get reduced form equation (VAR representation) such as:

$$\boldsymbol{y}_{ct} = \boldsymbol{A}_c^{-1}(\tau \boldsymbol{I_{N_c}} + \delta \boldsymbol{W}_c) \boldsymbol{y}_{ct-1} + \boldsymbol{A}_c^{-1} \boldsymbol{B}_c^{-1} \boldsymbol{v}_{ct} + \boldsymbol{A}_c^{-1} X_{ct}, \tag{5}$$

from which could be seen that this process is stable as long as following assumption is met:

Assumption 5. *For every country $c \in [1, ..., C]$, all eigenvalues of matrix $\boldsymbol{A}_c^{-1}(\tau \boldsymbol{I_{N_c}} + \delta \boldsymbol{W}_c)$ are in absolute value lesser than one.*

This assumption is then essential in context of estimation procedure described in following section, since for unstable and spatial cointegration cases that occur by validation of this assumption, different approaches are required to obtain consistent or more efficient estimation (described e. g. in Lee and Yu (2010) [6]).
Now lets look at the structure and features of explanatory variables and corresponding response parameters, which form the observed part of term X_{ct} in equations 4 and 5. Generally, three sets of explanatory variables are assumed - $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$. Within first set ($\boldsymbol{X_1}$), exogenous variables (potentially alongside with their spatial lag form) at regional level (i. e. there exist variation across all examined regions within these variables) are included, while the second set ($\boldsymbol{X_2}$) is made up of exogenous country-level variables (i. g. variables, that vary over countries but

have the same observed values for every region within the same country). Finally, endogenous variables could be divided into two parts: first part, which contains temporally and spatial-temporally lagged dependent variable, and second part expressed by set $\boldsymbol{X_3}$ which is made up of other endogenous variables which could be defined on regional as well as country level. For sets $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$ lets define corresponding response parameters as $\boldsymbol{\beta} = [\boldsymbol{\beta_1}^\intercal, \boldsymbol{\beta_2}^\intercal, \boldsymbol{\beta_3}^\intercal]^\intercal$. Important feature of this model is, that it allows for spatial heterogeneity at country level in response parameters of exogenous regional-level variables ($\boldsymbol{\beta_1}$) via random component within these parameters. More specifically, response parameters $\boldsymbol{\beta_1}$ are modelled as random variables which could be expressed as follows:

$$\boldsymbol{\beta}_{\mathbf{1},c} = \boldsymbol{\beta_1} + \boldsymbol{\epsilon}_c,\ E(\boldsymbol{\epsilon}) = \mathbf{0},\ Var(\boldsymbol{\epsilon}) = \boldsymbol{V}, \tag{6}$$

where matrix $\boldsymbol{V}$ is some positive definite symmetric matrix of dimension k_1 x k_1, where k_1 is number of variables contained in $\boldsymbol{X_1}$. This formulation means, that response parameters of exogenous regional-level variables could vary across countries around common means ($\boldsymbol{\beta_1}$).

To account for unobserved heterogeneous spatial means among dependent variable (denoted by $\boldsymbol{\eta}$), that could be possibly correlated to some of explanatory variables, one should subtract these means by some suitable transformation of the data (i. e. fixed effects (FE) specification).[1] For that purpose, Lee and Yu (2010) [5] propose transformation approach in following way:

$$[\boldsymbol{y}^*, \boldsymbol{X}^*] = (\boldsymbol{I_N} \otimes \boldsymbol{F}_{\boldsymbol{T-1}}^\intercal)[\boldsymbol{y}, \boldsymbol{X}], \tag{7}$$

where $\boldsymbol{F_{T-1}}$ contains eigenvectors of matrix $\boldsymbol{J_T} = \boldsymbol{I_T} - \boldsymbol{E_T}/T$ corresponding to eigenvalues equal to one (that is T x $T-1$ matrix).[2] Unlike the time-demeaning approach, transformation via eigenvectors matrix does not cause negative correlations across observations and further reduces dimension of the data from NT to $N(T-1)$, which could be useful in context of estimation procedures based on logarithmic likelihood (LL) function.[3]

By defining $\boldsymbol{K}_c = (\boldsymbol{\tau I_{N_c}} + \boldsymbol{\delta W}_c) \otimes \boldsymbol{I_{T-1}}$, complete model with all aforementioned features is formally expressed as:

$$\begin{aligned}
&\begin{bmatrix} \boldsymbol{A}_1 \otimes \boldsymbol{I_{T-1}} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_2 \otimes \boldsymbol{I_{T-1}} & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \boldsymbol{A}_C \otimes \boldsymbol{I_{T-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1^* \\ \boldsymbol{y}_2^* \\ . \\ \boldsymbol{y}_C^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{K}_1 & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K}_2 & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \boldsymbol{K}_C \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_{1t-1}^* \\ \boldsymbol{y}_{2t-1}^* \\ . \\ \boldsymbol{y}_{Ct-1}^* \end{bmatrix} + \\
&+ \begin{bmatrix} \boldsymbol{X}_{\mathbf{1},1}^* \\ \boldsymbol{X}_{\mathbf{1},2}^* \\ . \\ \boldsymbol{X}_{\mathbf{1},C}^* \end{bmatrix} \boldsymbol{\beta_1} + \begin{bmatrix} \boldsymbol{X}_{\mathbf{2},1}^* \\ \boldsymbol{X}_{\mathbf{2},2}^* \\ . \\ \boldsymbol{X}_{\mathbf{2},C}^* \end{bmatrix} \boldsymbol{\beta_2} + \begin{bmatrix} \boldsymbol{X}_{\mathbf{3},1}^* \\ \boldsymbol{X}_{\mathbf{3},2}^* \\ . \\ \boldsymbol{X}_{\mathbf{3},C}^* \end{bmatrix} \boldsymbol{\beta_3} + \begin{bmatrix} \boldsymbol{X}_{\mathbf{1},1}^* & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_{\mathbf{1},2}^* & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \boldsymbol{X}_{\mathbf{1},C}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ . \\ \boldsymbol{\epsilon}_C \end{bmatrix} + \\
&+ \begin{bmatrix} \boldsymbol{B}_1^{-1} \otimes \boldsymbol{I_{T-1}} & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_2^{-1} \otimes \boldsymbol{I_{T-1}} & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \boldsymbol{B}_C^{-1} \otimes \boldsymbol{I_{T-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ . \\ \boldsymbol{v}_C \end{bmatrix},
\end{aligned} \tag{8}$$

where country-specific component $\boldsymbol{v}_c$ is defined by following assumption:

Assumption 6. *The disturbances* $\{v_{cit}\}$ *are IID across* $i \in [1, .., N_c]$ *and* $t \in [1, .., T]$ *with zero mean, variance* $\sigma_{v_c}^2$ *and* $E|v_{cit}|^{4+\epsilon} < \infty$ *for some* $\epsilon > 0$.

Note that this assumption allows for heteroscedastic variances ($\sigma_{v_c}^2$) across countries and does not require normality. Complete residual (unobserved) component of model 8 is then given by:

$$\boldsymbol{u}^* = (\boldsymbol{A_N} \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}^* - \tilde{\boldsymbol{X}}\boldsymbol{\xi} = \boldsymbol{X}_{\mathbf{1},\boldsymbol{bd}}^*\boldsymbol{\epsilon} + (\boldsymbol{B_N} \otimes \boldsymbol{I_{T-1}})\boldsymbol{v}, \tag{9}$$

where $\boldsymbol{A_N}$ is block-diagonal matrix with C blocks where c-th block is given by matrix $\boldsymbol{A}_c$, $\boldsymbol{B_N}$ is defined in same way as $\boldsymbol{A_N}$, $\boldsymbol{X}_{\mathbf{1},\boldsymbol{bd}}^*$ is block-diagonal matrix with C blocks where c-th block is given by matrix $\boldsymbol{X}_{\mathbf{1},c}^*$ and for $\tilde{\boldsymbol{X}} = [\boldsymbol{y}_{\boldsymbol{t-1}}^*,\ (\boldsymbol{W_N} \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}_{\boldsymbol{t-1}}^*,\ \boldsymbol{X}_{\mathbf{1}}^*,\ \boldsymbol{X}_{\mathbf{2}}^*,\ \boldsymbol{X}_{\mathbf{3}}^*]$ applies following assumption:

Assumption 7. *The elements of* $\tilde{\boldsymbol{X}}$ *have finite first two moments, such as* $E(\tilde{x}_{k,it}^2) < \infty$ *for* $i \in [1, .., N]$, $t \in [1, .., T]$ *and* $k \in [1, .., k_{\tilde{x}}]$. *Also,* $\lim_{N\to\infty} \frac{1}{NT} \sum_i \tilde{\boldsymbol{X}}_i^\intercal \tilde{\boldsymbol{X}}_i$ *exists and is nonsingular with rank* $k_{\tilde{x}}$.

[1] Since $\boldsymbol{\eta}$ is subtracted, there is no need for formal description of its distribution. Only meaningful assumption about $\boldsymbol{\eta}$ is absence of infinite values, which is implicitly provided from assumptions 5 and 7.

[2] Note that matrix $\boldsymbol{J_T}$ is used when subtracting means via traditional time-demeaning.

[3] To deal with possible unobserved time effects, similar procedure is not plausible in context of model described in this article. Since time dimension is assumed as fixed while cross-sectional dimension tends to infinity, time dummy variables are sufficient estimators of these effects.

Corresponding variance matrix of unobserved component 9 ($\boldsymbol{\Sigma_N}$) then takes form of block-diagonal matrix with C blocks where c-th block given by:

$$\boldsymbol{\Sigma}_c = \boldsymbol{X}^*_{\mathbf{1},c}\boldsymbol{V}\boldsymbol{X}^{*\intercal}_{\mathbf{1},c} + \sigma^2_{v_c}(\boldsymbol{B}_c^\intercal\boldsymbol{B}_c \otimes \boldsymbol{I_{T-1}})^{-1}, \tag{10}$$

which is the same form as in model without endogenous spatial effects proposed by Elhorst and Zeilstra (2007) [4]. To sum up, model described by equation 8 contains country-specific parameters $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma^2_v} = [\sigma^2_{v,1}, ..., \sigma^2_{v,C}]^\intercal$, that is C x 3 country-specific parameters in total. Another group of parameters appearing in model 8 are response parameters, which are denoted by $\boldsymbol{\xi} = [\tau, \delta, \boldsymbol{\beta}^\intercal]^\intercal$. Remaining parameters are elements of matrix $\boldsymbol{V}$ and since this matrix is symmetric, we could take to account only $\frac{(k_1+1)k_1}{2}$ parameters (diagonal elements and elements above diagonal of $\boldsymbol{V}$). Therefore, there are C x 3 + $k_{\tilde{x}}$ + $\frac{(k_1+1)k_1}{2}$ parameters included in the model 8.

3 Estimation Procedure

To obtain consistent estimation of described model, analogue to Swamy's three stage procedure (Swamy (1970) [9]) taken and extended from Elhorst and Zeiltra (2007) [4] is applied. This section describes this procedure in sense of computation and provides some basic asymptotic properties (with $N \to \infty$) of obtained estimates, although rigorous asymptotic theory is not properly addressed in available literature and could be subject of future analysis. First stage of this method serves for two purposes. First, country-specific parameters $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma^2_v}$ are estimated and second, estimation of response parameters is performed in such way that we assume country-specific responses to exogenous regional-level variables (represented by country-specific response parameters $\boldsymbol{\beta_1} = [\boldsymbol{\beta}^\intercal_{\mathbf{1},1}, ..., \boldsymbol{\beta}^\intercal_{\mathbf{1},C}]^\intercal$). In other words, in first step we estimate model 8, but without random component $\boldsymbol{\epsilon}$ and with block diagonal matrix of explanatory variables contained in set $\boldsymbol{X_1}$, where c-th block is given by variables observed within c-th country. For estimation of parameters within this set up, it is applied procedure that alternates between maximizing country-specific concentrated LL functions to obtain estimates of spatial parameters ($\boldsymbol{\rho}$, $\boldsymbol{\lambda}$) and IVFGLS method for $\boldsymbol{\sigma^2_v}$ and response parameters $\boldsymbol{\beta}$ estimation. First, it is useful to express LL function that corresponds to the model 8 with restriction $\boldsymbol{V} = \boldsymbol{0}$ and matrix with transformed exogenous regional variables $\boldsymbol{X^*_{1,bd}}$ characterized by aforementioned block diagonal structure. By using Jacobian of transformation and applying general formula of LL function as if disturbances $\boldsymbol{\nu}$ were normally distributed (as described e. g. in Anselin (1987) [2]), this function takes form:

$$\begin{aligned} LL(\boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\sigma^2_v}, \boldsymbol{\xi}|\boldsymbol{\eta}) = -\tfrac{N(T-1)}{2}\log(2\pi) + (T-1)\sum_c \log|\boldsymbol{A}_c| + \\ +(T-1)\sum_c \log|\boldsymbol{B}_c| - \sum_c \frac{N_c(T-1)}{2}\log(\sigma^2_{v,c}) - \frac{1}{2}\boldsymbol{u}^{*\intercal}\boldsymbol{\Sigma}_{\boldsymbol{N}}^{-1}\boldsymbol{u}^*, \end{aligned} \tag{11}$$

where matrix $\boldsymbol{\Sigma_N}$ is again block diagonal matrix where c-th block is given by 10 with restriction such as $\boldsymbol{V = 0}$ and $\boldsymbol{u^*}$ is specified by 9. Because heterogeneous means ($\boldsymbol{\eta}$) are concentrated out via transformation described by 7, LL function is conditional on $\boldsymbol{\eta}$. Although maximization of this function does not yield consistent estimators as long as set of endogenous variables ($\boldsymbol{X_3}$) is nonempty, this formulation is useful because when we have consistent estimates for $\boldsymbol{\xi}$, we could use concentrated LL function (conditional on $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$) to obtain consistent spatial parameters. Using first order conditions for $\boldsymbol{\xi}$ and $\boldsymbol{\sigma^2_{v_c}}$, which are given by:

$$\boldsymbol{\xi} = (\tilde{\boldsymbol{X}}^\intercal(\boldsymbol{B}_{\boldsymbol{N}}^\intercal\boldsymbol{B}_{\boldsymbol{N}} \otimes \boldsymbol{I_{T-1}})\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\intercal(\boldsymbol{B}_{\boldsymbol{N}}^\intercal\boldsymbol{B}_{\boldsymbol{N}} \otimes \boldsymbol{I_{T-1}})(\boldsymbol{A}_{\boldsymbol{N}} \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}^* \tag{12}$$

$$\sigma^2_{v_c} = \frac{\boldsymbol{u}^{*\intercal}_c(\boldsymbol{B}_c^\intercal\boldsymbol{B}_c \otimes \boldsymbol{I_{T-1}})\boldsymbol{u}^*_c}{N_c(T-1)}, \tag{13}$$

concentrated LL could be written as:

$$\begin{aligned} LL(\boldsymbol{\lambda}, \boldsymbol{\rho}|\boldsymbol{\sigma^2_v}, \boldsymbol{\xi}, \boldsymbol{\eta}) = Cn + (T-1)\sum_c \log|\boldsymbol{A}_c| + (T-1)\sum_c \log|\boldsymbol{B}_c| - \\ -\sum_c \frac{N_c(T-1)}{2}\log[\boldsymbol{u}^{*\intercal}_c(\boldsymbol{B}_c^\intercal\boldsymbol{B}_c \otimes \boldsymbol{I_{T-1}})\boldsymbol{u}^*_c], \end{aligned} \tag{14}$$

where Cn is a constant which does not depend on any of the parameters. This function has desirable feature such as it could be rewritten to additive form:

$$LL(\boldsymbol{\theta}) = \sum_c LL_c(\boldsymbol{\theta}_c), \tag{15}$$

with

$$LL_c(\boldsymbol{\theta}_c) = Cn + (T-1)\log|\boldsymbol{A}_c| + (T-1)\log|\boldsymbol{B}_c| - \frac{N_c(T-1)}{2}\log[\boldsymbol{u}^{*\intercal}_c(\boldsymbol{B}_c^\intercal\boldsymbol{B}_c \otimes \boldsymbol{I_{T-1}})\boldsymbol{u}^*_c], \tag{16}$$

where $\boldsymbol{\theta}_c$ contains every parameter appearing in the c-th concentrated LL function. This future is important, since it implies that separate maximization of country-specific LL functions yields the same estimates as maximization of full LL function described by 14. This allows for easier optimization, since for each $LL_c(\boldsymbol{\theta}_c)$ only two parameters (λ_c, ρ_c) are estimated at the same time. Furthermore, analogically to simple SAR model (see e. g. LeSage and Pace (2009) [7]), each country-specific concentrated LL function (defined by 16) could be rewritten to following form:

$$\begin{aligned} LL_c(\boldsymbol{\theta}_c) = Cn - \tfrac{N(T-1)}{2} \log[(\boldsymbol{e_{0,c}} - \lambda_c \boldsymbol{e_{1,c}})^\top (\boldsymbol{B}_c^\top \boldsymbol{B}_c \otimes \boldsymbol{I_{T-1}})(\boldsymbol{e_{0,c}} - \lambda_c \boldsymbol{e_{1,c}})] + \\ +(T-1)\log|\boldsymbol{A}_c| + (T-1)\log|\boldsymbol{B}_c|, \end{aligned} \tag{17}$$

with $\boldsymbol{e_0}$ and $\boldsymbol{e_1}$ defined as:

$$\boldsymbol{e_0} = \boldsymbol{y}^* - \tilde{\boldsymbol{X}}\boldsymbol{\xi_d} \tag{18}$$

$$\boldsymbol{e_1} = (\boldsymbol{W}_N \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}^* - \tilde{\boldsymbol{X}}\boldsymbol{\xi_w} \tag{19}$$

while between terms $\boldsymbol{\xi_d}$, $\boldsymbol{\xi_w}$ and $\boldsymbol{\xi}$ exist relationship such as:

$$\tilde{\boldsymbol{X}}\boldsymbol{\xi} = \tilde{\boldsymbol{X}}\boldsymbol{\xi_d} - \boldsymbol{D}_\lambda \tilde{\boldsymbol{X}}\boldsymbol{\xi_w}, \tag{20}$$

where $\boldsymbol{D}_\lambda$ is diagonal matrix with C blocks and c-th block given by $\lambda_c \boldsymbol{I}_{N_c(\boldsymbol{T-1})}$. From this formulation it could be seen, that consistent estimates of $\boldsymbol{\xi_d}$ and $\boldsymbol{\xi_w}$ are sufficient for construction of the concentrated LL function. Because among explanatory variables are those which are assumed to be endogenous, for consistent estimates of aforementioned expressions are required techniques using instrumental variables. Furthermore, as shown in Alvarez and Arellano (2003) [1], asymptotic bias of autoregressive coefficient within simple AR(1) model in panel data set up with fixed cross-sectional effects is of order $O(\frac{1}{T})$ (thus inconsistent for T fixed), while e. g. for GMM estimators proposed by Arellano and Bond (1991) [3] (that use IV estimation), there exists asymptotic bias of order $O(\frac{1}{N})$, which is much more suitable in context of model with short time dimension relatively to the cross-sectional dimension. For purposes of further analysis, following assumptions about instruments and initial observations are made:

Assumption 8. *Variables and instruments contained in matrix* $\boldsymbol{Z}$ *are exogenous such as* $E(\boldsymbol{u}|\boldsymbol{Z}) = \boldsymbol{0}$.

Assumption 9. $E(\boldsymbol{Z}^\top \boldsymbol{Z})$ *has rank* k_z *and* $E(\boldsymbol{Z}^\top \tilde{\boldsymbol{X}})$ *has rank* $k_{\tilde{x}}$*, where* $k_z \geq k_{\tilde{x}}$ *is number of columns of* $\boldsymbol{Z}$.

Assumption 10. *Initial time observations* $\boldsymbol{y}_{i0}$ *are assumed to be exogenous.*

Based on these assumptions, estimation of $\boldsymbol{e_0}$ and $\boldsymbol{e_1}$ (respectively $\boldsymbol{\xi_d}$ and $\boldsymbol{\xi_w}$) takes form:

$$\hat{\boldsymbol{e}}_{\boldsymbol{0}} = (\boldsymbol{I}_{N(\boldsymbol{T-1})} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top \boldsymbol{P}_{\boldsymbol{Z}} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{y}^* = \boldsymbol{y}^* - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\xi}}_{\boldsymbol{d}} \tag{21}$$

$$\hat{\boldsymbol{e}}_{\boldsymbol{1}} = [\boldsymbol{I}_{N(\boldsymbol{T-1})} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top \boldsymbol{P}_{\boldsymbol{Z}} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{P}_{\boldsymbol{Z}} (\boldsymbol{W}_N \otimes \boldsymbol{I_{T-1}})]\boldsymbol{y}^* = (\boldsymbol{W}_N \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}^* - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\xi}}_{\boldsymbol{w}}, \tag{22}$$

where $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top$ is projection matrix of $\boldsymbol{Z}$. These estimates are then proven to be consistent as long as assumptions 1, 5, 6, 7, 8, 9 and 10 hold (these conditions are sufficient, not necessary).[4] Provided that consistent estimates of $\boldsymbol{\xi_d}$ and $\boldsymbol{\xi_w}$ are available, by maximizing concentrated LL functions defined by 17 we could obtain consistent estimates of spatial parameters $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}$ provided that assumptions 1, 4, 5 and 6 hold.[5] Next, by using estimated spatial parameters $\hat{\rho}$, $\hat{\lambda}$ and expressions $\hat{\boldsymbol{\xi}}_{\boldsymbol{d}}$, $\hat{\boldsymbol{\xi}}_{\boldsymbol{w}}$ from previous steps, estimates of country-specific variances are obtained from expression 13 using relationships described by 18, 19 and 20. Further, estimation of $\boldsymbol{\xi}$ is obtained via IVFGLS formula given by:

$$\hat{\boldsymbol{\xi}} = (\tilde{\boldsymbol{X}}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z}(\boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z}(\boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}_N^{-1} (\hat{\boldsymbol{A}}_N \otimes \boldsymbol{I_{T-1}})\boldsymbol{y}^*, \tag{23}$$

where, as mentioned earlier, $\hat{\boldsymbol{\Sigma}}_N$ is within this stage of Swamy's procedure assumed with restriction such as $\boldsymbol{V} = \boldsymbol{0}$. If exogenous spatial effects are not included ($\boldsymbol{\rho} = \boldsymbol{0}$), then first stage of estimation ends. If it is not the case, estimate of $\boldsymbol{\xi}$ is further used in construction of concentrated LL function 16, which is then again maximized to obtain improved estimates of spatial parameters. These steps, which serve to improve efficiency of the estimates, are repeated until convergence occurs.

Estimated country-specific response parameters ($\hat{\boldsymbol{\beta}}_{\boldsymbol{1},c}$) are then used to obtain estimate of matrix $\boldsymbol{V}$ within second stage of the estimation procedure. According to Elhorst and Zeilstra (2007) [4], this matrix could be consistently estimated as:

$$\hat{\boldsymbol{V}} = \frac{1}{C-1} \sum_c (\hat{\boldsymbol{\beta}}_{\boldsymbol{1},c} - \frac{1}{C} \sum_c \hat{\boldsymbol{\beta}}_{\boldsymbol{1},c})((\hat{\boldsymbol{\beta}}_{\boldsymbol{1},c} - \frac{1}{C} \sum_c \hat{\boldsymbol{\beta}}_{\boldsymbol{1},c})^\top. \tag{24}$$

[4] These conditions are derived from standard asymptotic theory for 2SLS method, see e. g. Wooldridge (2002) [10] for more detail.

[5] These conditions are implied from analysis made by Lee and Yu (2010) [5] for non dynamic specification with both endogenous and exogenous spatial effects.

Within final, third stage of this procedure, common parameters $\boldsymbol{\beta_1}$ alongside with other parameters contained in $\boldsymbol{\xi}$ corresponding to model 8 are estimated via IVFGLS formula 23, where estimates of $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$, σ_v^2 and $\boldsymbol{V}$ from previous stages are used to form expressions $\boldsymbol{\Sigma_N}$ and $\boldsymbol{A_N}$. Variance matrix of $\boldsymbol{\xi}$ is then derived from IVFGLS procedure and could be consistently estimated by:

$$\hat{Var}(\hat{\boldsymbol{\xi}}) = (\tilde{\boldsymbol{X}}^\intercal \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z} (\boldsymbol{Z}^\intercal \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}^\intercal \hat{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{X}})^{-1}. \tag{25}$$

Standard errors of spatial parameters λ_c and ρ_c are then obtained from inverse of the Hessian matrices of corresponding concentrated LL functions.[6] Main disadvantage of this methodology is, that variances of spatial parameters and variance matrix of $\boldsymbol{\xi}$ are estimated separately, thus there is no estimated covariance between them. This could be a slight problem in case of spatial parameters $\boldsymbol{\lambda}$, since unlike in case of parameters $\boldsymbol{\rho}$, some form of covariance with parameters $\boldsymbol{\xi}$ is expected. Main reason for interest in this covariance is computation of inference of direct, indirect and total effects. Because of this limitation, restrictions such as $Cov(\lambda_c, \lambda_r) = 0$, $Cov(\rho_c, \rho_r) = 0,\ \forall\ c, r \neq c \in [1, ..., C]$ and $Cov(\boldsymbol{\xi}, \lambda_c) = \boldsymbol{0}$, $Cov(\boldsymbol{\xi}, \rho_c) = \boldsymbol{0},\ \forall\ c \in [1, ..., C]$ are imposed.[7]

4 Conclusions

This theoretically oriented article combines findings, methodologies and conclusions especially from work of Elhorst and Zeilstra (2007) [4] and Lee and Yu (2010) [5] to define complex multilevel spatial dynamic model with heterogeneous structure and possibly endogenous regressors and provide procedure for estimation of corresponding parameters. Important extension compared to model proposed by Elhorst and Zeilstra (2007) [4] is implementation of endogenous spatial effects, which are represented by SAR process that is heterogeneous across first level of the cross-sectional dimension. Estimation procedure then uses adjusted Swamy's three step procedure (Swamy (1970) [9]) which, based on conclusions from asymptotic properties of some nested forms models, provide consistent estimators.

Acknowledgements

The work was supported by the Prague University of Economics and Business project IGS F4/24/2023.

References

[1] Alvarez, J. & Arellano, M. (2003) The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica* 71.4, pp. 1121–1159.

[2] Anselin, L. (1988) *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

[3] Arellano, M. & Bond, S. (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies* 58.2, pp. 277–297.

[4] Elhorst, J. P. & Zeilstra, A. S. (2007) Labour force participation rates at the regional and national levels of the European Union: An integrated analysis. *Papers in Regional Science* 86.4, pp. 525–549.

[5] Lee, L. & Yu, J. (2010a) Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154.2, pp. 165–185.

[6] Lee, L. & Yu, J. (2010b) Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40.5. Advances In Spatial Econometrics, pp. 255–271.

[7] Lesage, J. P. & Pace, R. K. (2009) *Introduction to spatial econometrics.* CRC Press.

[8] Millo, G. & Piras, G. (2012) splm: Spatial Panel Data Models in R. *Journal of Statistical Software* 47.1, pp. 1–38.

[9] Swamy, P. A. V. B. (1970) Efficient Inference in a Random Coefficient Regression Model. *Econometrica* 38.2, pp. 311–323.

[10] Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data.* Cambridge and London: MIT Press.

[6] These estimates of standard errors are consistent by making additional normality assumption about disturbances $\boldsymbol{v}$ (see e. g. Lee and Yu (2010) [5]).

[7] Similar way to obtain inference for parameters in model with endogenous spatial effects defined by SAR process is described in Millo and Piras (2012) [8], where they use similar iterative procedure for estimation of SAR model with random cross-sectional means.

Determining the Optimal Location of the Logistics Center in the Presence of Limiting Conditions

Josef Košťálek[1], Pavla Koťátková Stránská[2]

Abstract. The problem of the optimal location of a logistics center that is supposed to supply, for example, shops is often solved in logistics. Optimum location is important for saving transport costs. The problem is, however, in the existence of limiting conditions that generate areas where a logistics center cannot be built.
This article describes a mathematical model that is able to solve this problem and interpret the results graphically. The sum of transportation costs is an objective function, the mathematical model calculates the coordinates of the logistics center so that the value of the objective function is minimal. The existence of restricted areas creates restrictive conditions that are part of this model. Restricted areas can be defined using various mathematical shapes (circle, ellipse, rectangle, etc.). From a mathematical point of view, this situation is a deployment problem, but modified by restricted areas. And to solve the problem defined in this way, limiting conditions in the form of implication were used. If the x-coordinate of the logistics center lies in a restricted area, the value of the y-coordinate must be such that the logistics center is located outside the restricted area. For example when the restricted area has an ellipse shape the constraint condition for the y coordinate must work with the equation of the ellipse. Of course, this problem does not arise only when building a logistics center, but in every situation where it is necessary to determine the correct position of a central point that is inserted into a set of points. And there are links between the points and the central point. Finding the optimal location to place: a new machine in a factory, a waste dump, a grain silo, etc.

Keywords: mathematical model, limiting conditions, modified deployment problem

JEL Classification: C44
AMS Classification: 90C30

1 Introduction

Our article describes a mathematical solution to the placement problem when we are looking for a suitable location for, for example, a logistics center. The logistics center will have coordinates [X, Y], the points it can supply (e.g. shops) will have coordinates [x_i, y_i]. But the amount of supplies transported to each point (symbol Q_i) will be different (the larger the amount, the thicker the line) see Figure 1.

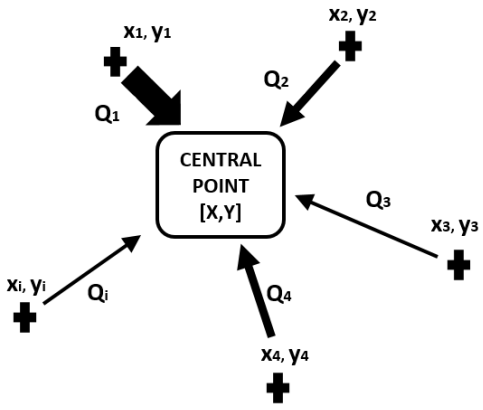


Figure 1 Description of the situation
Source: own

[1] University of Chemistry and Technology, Prague, School of Business, Technická 5, 166 28 Prague 6, josef.kostalek@vscht.cz.
[2] University of Chemistry and Technology, Prague, School of Business, Technická 5, 166 28 Prague 6, pavla.kotatkova.stranska@vscht.cz.

1.1 The Existence of Restricted Areas

Solving this problem is not difficult, see formula 1. The result is the [X, Y] coordinates determining the location of, for example, a logistics center.

$$F = \sum_{i=1}^{n} \sqrt{(X - x_i)^2 + (Y - y_i)^2} \cdot Q_i = min. \tag{1}$$

The logistics center cannot be placed, for example, in mountains, in forests, in city centers, etc. Therefore, we define restricted areas of shapes: circles, rectangles and ellipses. Then we need to add restrictive conditions for the [X, Y] coordinates to the mathematical model. Moreover, the rectangle or ellipse does not have to be only in a horizontal position, but it is possible to define their rotation. Figure 2 describes the situation.

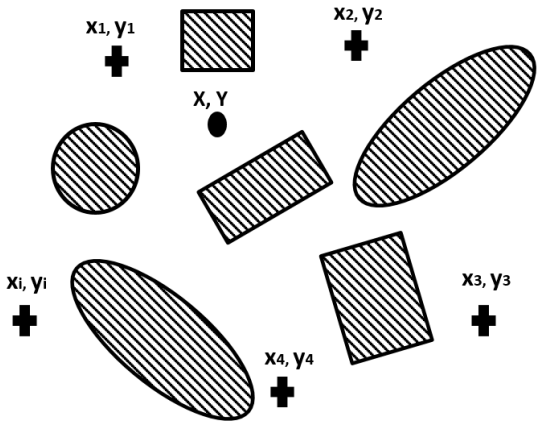


Figure 2 Restricted areas of shapes: circles, rectangles and ellipses
Source: own

The mathematical model will contain an objective function (formula 1) and restrictive conditions that ensure that the central point does not situate in the restrictive areas. By solving the mathematical model, the resulting coordinates [X, Y] will be such that the objective function will be minimal and the restrictive conditions will be fulfilled. This article describes a mathematical model that uses an exact solution. But there are other options for layout problems, namely solutions using heuristic or genetic algorithms (Sun et al. [3], Syam [4], Zhou et al. [5]).

2 Restrictive Conditions for Restricted Areas

2.1 Restrictive Conditions for Restricted Areas of Circle Shapes

In this case, we will use restrictive conditions in the form of implication. Formulas 2, 3, 4, 5 generally describe the situation (Jablonský [1]).

$$(A \Rightarrow B) \Leftrightarrow (\neg A \vee B) \tag{2}$$

For example:

$$IF\ x_1 < 2\ THEN\ x_2 \leq 3 \iff x_1 \geq 2\ OR\ x_2 \leq 3 \tag{3}$$

The mathematical notation of such conditions is described by formulas 4 and 5.

$$x_1 \geq 2 - M \cdot t \tag{4}$$

$$x_2 \leq 3 - M \cdot (1 - t), where\ t\ \epsilon\ \{0; 1\} \tag{5}$$

Where: x_1 and x_2 are variables, $t \in \{0;1\}$ (binary variable) and M is a large number (for example $M = 10^6$).

This exact procedure can be applied to the mathematical notation of the restrictive conditions that arise when there is a restricted area in the shape of a circle, (Koťátková Stránská, Košťálek [2]) see Figure 3 and formulas 6, 7, 8.

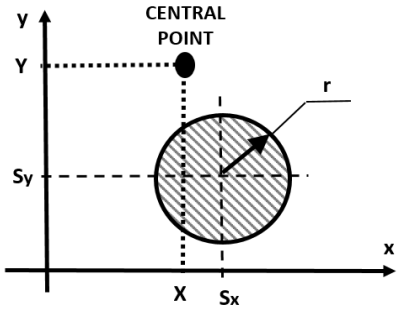


Figure 3 Constraints for central point coordinates [X, Y] for restricted area circle shapes
Source: own

$$IF(X-S_x)^2 \le r^2\ THEN\ \left(Y-S_y\right)^2 \ge r^2-(X-S_x)^2 \Leftrightarrow\ (X-S_x)^2 > r^2 OR\ \left(Y-S_y\right)^2 \ge r^2-(X-S_x)^2 \quad (6)$$

$$(X-S_x)^2 > r^2 - M\cdot t \quad (7)$$

$$\left(Y-S_y\right)^2 \ge r^2-(X-S_x)^2 - M\cdot(1-t) \quad (8)$$

Where: t ϵ {0;1} (binary variable), M = 10^6 (constant)

2.2 Restrictive Conditions for Restricted Areas of Ellipse Shapes

Here it was necessary to solve the situation where the ellipse is rotated by an angle β. But in such a situation, the principle used for the circle would not work. We used a solution where we rotate the entire coordinate system by an angle minus β. Then the ellipse is rotated to a horizontal position and we can use the same method as for the circle. But it is necessary to transform the coordinates of the central point [X, Y] to [X_{TR}, Y_{TR}], the coordinates transformed in this way are inserted into the restrictive conditions, see Figure 4 and formulas 9, 10, 11, 12, 13.

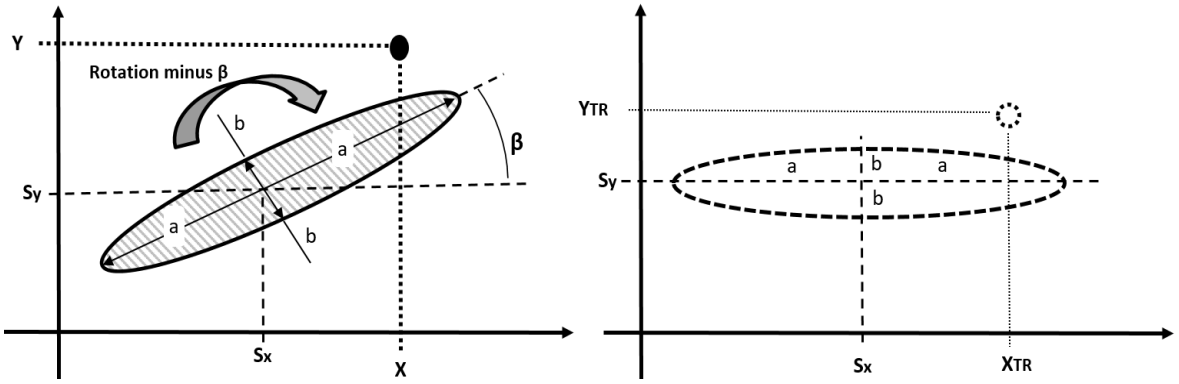


Figure 4 Constraints for central point coordinates [X, Y] for restricted area ellipse shapes
Source: own

Transformation [X, Y] to [X_{TR}, Y_{TR}], where the axis of rotation is not at a point [0, 0], but a point [S_x, S_y]:

$$X_{TR} = S_x + (X-S_x)\cdot\cos(-\beta) - (Y-S_y)\cdot\sin(-\beta) \quad (9)$$

$$Y_{TR} = S_y + (X-S_x)\cdot\sin(-\beta) + (Y-S_y)\cdot cos(-\beta) \quad (10)$$

Condition for ellipse in horizontal position:

$$IF(X_{TR}-S_x)^2 \le a^2\ THEN\ \left(Y-S_y\right)^2 \ge \left[1-\frac{(X_{TR}-S_x)^2}{a^2}\right]\cdot b^2 \Leftrightarrow (X_{TR}-S_x)^2 > a^2\ OR\ \left(Y_{TR}-S_y\right)^2 \ge \left[1-\frac{(X_{TR}-S_x)^2}{a^2}\right]\cdot b^2 \quad (11)$$

Restrictive conditions in the mathematical model:

$$(X_{TR}-S_x)^2 > a^2 - M\cdot t \quad (12)$$

$$\left(Y_{TR}-S_y\right)^2 \ge \left[1-\frac{(X_{TR}-S_x)^2}{a^2}\right]\cdot b^2 - M\cdot(1-t) \quad (13)$$

Where: t ϵ {0;1} (binary variable), M = 10^6 (constant)

2.3 Restrictive Conditions for Restricted Areas of Rectangle Shapes

The restricted area is a rectangle that can be rotated by an angle α. In this situation, it is not possible to use the same method as for the situation where the restrictive area is in the shape of an ellipse or a circle. The figure 5 describes the situation. The central point must be located outside the rectangle. Four half-planes are created outside the rectangle, which are mathematically written by formulas 14, 15, 16, 17.

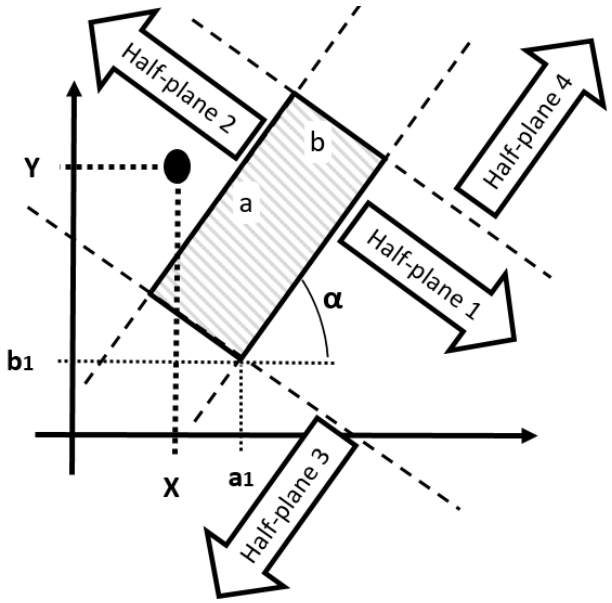


Figure 5 Constraints for central point coordinates [X, Y] for restricted area rectangle shapes
Source: own

Half-plane 1:

$$Y \leq b_1 + tg(\alpha) \cdot (X - a_1) \tag{14}$$

Half-plane 2:

$$Y \geq b_1 + \frac{b}{\cos \alpha} + tg(\alpha) \cdot (X - a_1) \tag{15}$$

Half-plane 3:

$$Y \leq b_1 - tg(90° - \alpha) \cdot (X - a_1) \tag{16}$$

Half-plane 4:

$$Y \geq b_1 + \frac{a}{\sin \alpha} - tg(90° - \alpha) \cdot (X - a_1) \tag{17}$$

The restrictive conditions in the mathematical model were created using the equations of the four half-planes from formulas 14, 15, 16 and 17. The restrictive conditions are written in formulas 18, 19, 20, 21 and 22. These conditions for the center point [X, Y] coordinates ensure that the central point is situated outside the rectangle.

$$Y \leq b_1 + tg(\alpha) \cdot (X - a_1) + t_1 \cdot M \tag{18}$$

$$Y \geq b_1 + \frac{b}{\cos \alpha} + tg(\alpha) \cdot (X - a_1) - t_2 \cdot M \tag{19}$$

$$Y \leq b_1 - tg(90° - \alpha) \cdot (X - a_1) + t_3 \cdot M \tag{20}$$

$$Y \geq b_1 + \frac{a}{\sin \alpha} - tg(90° - \alpha) \cdot (X - a_1) - t_4 \cdot M \tag{21}$$

$$t_1 + t_2 + t_3 + t_4 \leq 3 \tag{22}$$

Where: t ϵ {0;1} (binary variable), M = 10^6 (constant)

3 Graphical Output from the Model

The previous part of the article described the creation of a mathematical model containing an objective function and many restrictive conditions. This part of the article describes the functioning of the model when solving a specific problem, as well as methods for graphically drawing the situation, see Figure 6.

	INPUT 1st STEP			2nd STEP	OUTPUT		
	POINTS		INTENS.	Real. dist.	Correction	CENTRAL POINT	
	Xi	yi	Qi	ri	ki	X	Y
1	75	90	40		1	137,7	129,7
2	50	180	50		1		
3	150	230	35		1	Fmin.	43 182,34
4	200	30	40		1		
5	125	150	20		1		
6	200	190	10		1		
7	300	130	30		1		
8	250	80	20		1		
9	275	212	30		1		
10	450	120	20		1		
11	375	100	10		1		
12	70	135	32		1		
13	110	70	40		1		
14					1		
15					1		
16					1		
17					1		
18					1		
19					1		
20					1		

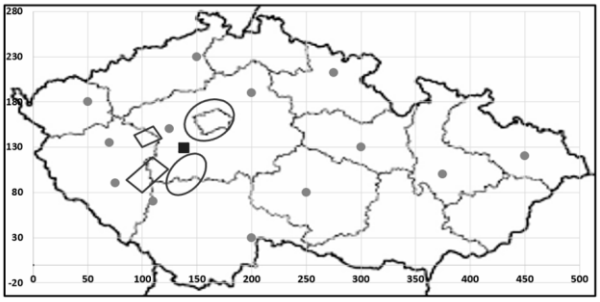


Figure 6 A view of the output from the model
Source: own

The coordinates of points and the intensity of transport are entered into the model, these points are, for example, shops. They are automatically drawn as small filled circles in the graph. Furthermore, the areas in which the central point cannot be placed are entered and these are also drawn in the graph (rectangles, circles, ellipses) with the possibility of rotation. The model can solve the problem of where to place the central point so that transportation costs are minimal. The central point is drawn as a small filled square. The central point is, for example, a logistics center that will supply shops. In this case, a map of the Czech Republic is inserted into the background of the graph. Points are plotted on the graph using x, y coordinates. The rectangle is drawn using four lines that connect four points (vertices of the rectangle A, B, C, D). Circle and ellipse are drawn as a curve connecting points with x, y coordinates which are calculated from polar coordinates.

To define the restricted area of the circle shape, the input are: S_x, S_y, r, see figure 3. Then the coordinates for plotting in the graph are calculated using formulas 23 and 24.

$$x = S_x + r \cdot cos\varphi,\ where\ \varphi\ \epsilon < 0; 2\pi > \qquad (23)$$

$$y = S_y + r \cdot sin\varphi,\ where\ \varphi\ \epsilon < 0; 2\pi > \qquad (24)$$

In order to make rotation possible for ellipses and rectangles, it is necessary to use general mathematical formulas to change the coordinates of a point [x, y] after rotation by an angle of α size to point [x´,y´] see formulas 25 and 26.

$$\acute{x} = x \cdot cos\ \alpha - y \cdot sin\alpha \qquad (25)$$

$$\acute{y} = x \cdot sin\alpha + y \cdot cos\alpha \qquad (26)$$

To define the restricted area of the ellipse shape, the input are: S_x, S_y, a, b, β see figure 4. Then the coordinates for plotting in the graph are calculated using formulas 27 and 28. These formulas are the equations of the ellipse in polar coordinates, which are transformed according to formulas 25 and 26 when rotated.

$$x = S_x + a \cdot cos\ (\varphi) \cdot cos\beta - b \cdot \sin(\varphi) \cdot sin\beta,\ where\ \varphi\ \epsilon < 0; 2\pi > \qquad (27)$$

$$y = S_y + a \cdot cos\ (\varphi) \cdot sin\beta + b \cdot \sin(\varphi) \cdot cos\beta,\ where\ \varphi\ \epsilon < 0; 2\pi > \qquad (28)$$

To define the restricted area of the rectangle shape, the input are: a_1, b_1, a, b, α see figure 5. To calculate the coordinates of vertices A, B, C, D after rotation, we will again use formulas 25 and 26 (the same principle).

4 Solving the Euclidean Metric Problem

In this phase, we described the principle of operation of the mathematical model capable of calculating [X, Y] coordinates, for which the value of the objective function will be minimal and the restrictive conditions will be observed. The problem is that the distances between the embedded points and the central point are straight-line distances (the Euclidean metric). But in a real situation, the distances are greater because the routes are on roads.

We will make one modification to the model described above, namely in the objective function, see formula 29.

$$F = \sum_{i=1}^{n} \sqrt{(X - x_i)^2 + (Y - y_i)^2} \cdot k_i \cdot Q_i = min.,\ where\ k_i = \frac{distance\ in\ reality}{distance\ in\ Euclidean\ metric} \qquad (29)$$

In the first step, a calculation is made where $k_1 = 1$, $k_2 = 1$, $k_3 = 1$ and we get the result [$X_{(0)}$, $Y_{(0)}$]. In the second step, we determine the k_i values for each point (for example: $k_1 = 2.3$, $k_2 = 1.4$ etc.), insert them into the objective

function and repeat the calculation, then we get the result $[X_{(1)}, Y_{(1)}]$. Step two is repeated as long as the differences between the results $[X_n, Y_n]$ and $[X_{n+1}, Y_{n+1}]$ are significant.

5 Conclusion

This article describes the way in which it is possible to solve the problem of optimal placement of the central point for a modified situation where we need to define areas in which the central point cannot be placed. Mathematical methods and formulas have been described that will allow the problem to be written using a mathematical model. In order for this mathematical model to be of practical use, it is necessary to use appropriate software to solve it. Various optimization software are available, but in this case it is also possible to use MS Excel, which has the advantage of being available, cheap and easy for users. The main goal of this article was to create a mathematical model, but it also describes a model in MS Excel, where it is possible to enter input values and, after starting the "solver" tool, in a short time find out the sought coordinates of the central point and a graphical drawing of the entire situation.

The input to the model is the coordinates of a set of points, the intensity of transport to these points from the central point and the definition of areas where the central point cannot be placed. The model allows you to insert a set of twenty points. And the restricted areas can be ten rectangles, ten circles and ten ellipses. Each area is defined by the size, position and angle of rotation, see Figures 3, 4, 5. The values are simply inserted into the prepared tables and all the calculations described in the article are performed automatically. All described restrictive conditions are set in the "solver" tool and it is ready to calculate the results of the mathematical model. Further calculations are performed automatically, where the results are the necessary coordinates and the entire situation (input, output and results) is drawn graphically.

The search for the optimal location of a logistics center was chosen as an example for the practical use of these calculations and this model. But we can solve the same problem in other situations, e.g. when placing a new machine among the original machines in a factory, when we are looking for a place for a waste dump, when we are looking for a place for a helipad or when placing a warehouse in intra-company logistics.

This mathematical model for the deployment problem contains a number of modifications to be as close as possible to the real situation (restrictive areas, repeating the calculation after entering the correction coefficients k_i). But it's still a model. The final decision on the location of the central point (e.g. logistics center) is in the hands of management, which must consider other criteria. The situation may look like this, with the help of the model, we determine that the optimal location of the logistics center is in the Central Bohemian region, 50 km west of Prague. According to the real situation, we place restrictive areas in this region (restrictive areas are also a place where municipalities do not allow the construction of a logistics center in their plans), calculate the values of k_i, which we insert into the objective function and repeat the calculation. We get a mathematical result. And we further specify it according to the situation, e.g. in the vicinity of the mathematical result [X, Y] we look for a place where it is possible to build and where there is an exit from the highway.

References

[1] Jablonský, J. (2007). *Programy pro matematické modelování*. Praha: VŠE. pp. 100-103. ISBN 978-80-245-1810-7.

[2] Koťátková Stránská, P. & Košťálek J. (2019). Solving a deployment problem with restricted areas. In 18th *Conference on Applied Mathematics* (*APLIMAT)*. Bratislava: Slovak University of Technology. pp. 685-690. ISBN 978-1-5108-8214.

[3] Sun, H., Gao, Z. & Wu, J. (2008). *A bi-level programming model and solution algorithm for the location logistics distribution centers*. Applied Mathematical Modelling, 32(4), 610-616. https://doi.org/10.1016/j.apm.2007.02.007.

[4] Syam, S. S. (2002). *A model and methodologies for the location problem with logistical components.* Com puters & Operations Research, 29(9). 1173-1193. https://doi.org/10.1016/S0305-0548(01)00023-5.

[5] Zhou, G., Min, H. & Gen, M. (2002). *The balanced allocation of customers to multiple distribution centers in the supply chain network: a genetic algorithm approach.* Computers & Industrial Engineering, 43(1), 251–261. https://doi.org/10.1016/S0360-8352(02)00067-0.

Development of the Efficiency of the Czech Automotive Industry

Richard Kovárník[1], Michaela Staňková[2]

Abstract. The automotive industry is an integral part of every European economy. In the Czech Republic, this sector is one of the most important industrial sectors. This article deals with an evaluation of production efficiency of the Czech automotive industry with regard to the size of the enterprises. For the purpose of evaluating efficiency, this article used the method of data envelopment analysis. Calculations are based on accounting data of individual enterprises with available data for the period between 2010 and 2019. Radial input-oriented BCC models are constructed separately for individual years. Efficiency is derived based on the amount of capital and the cost of employees in contrast to total amount of sales and the profit and loss statement in individual years. In addition, the Malmquist index was used to calculate changes in efficiency over time. The results of our analysis show that differentiating efficiency results with respect to the size of the enterprise makes sense, as especially small enterprises differ considerably in their results.

Keywords: Czech Republic, data envelopment analysis, efficiency, automotive industry

JEL Classification: C44, D24
AMS Classification: 90B50, 90C08

1 Introduction

During the 20th century, European car manufacturers became some of the world's most important producers. The success of local companies and their economic importance contributed to economic development and the orientation of economic policy. The automotive sector has become a major source of innovation as well as a target for investment in science and research. The European automotive industry is the source of up to a third of global production, employing over 12 million people [1]. Countries that are strongly historically oriented include Germany, France, Italy and, to a large extent, the Czech Republic as well. The dramatic changes in the last decade and the planned transformation of the sector have forced producers to think about innovation and the maximum use of resources so that their consumption is as efficient as possible [2].

This study is focused on evaluating the efficiency of the automotive industry in the Czech Republic. The Czech Republic is a country that is strongly oriented towards the automotive sector. Vehicle production in the Czech Republic accounts for approximately a quarter of total industrial production, a quarter of exports, and employs over 400,000 people [3]. There is no doubt about the economic importance of the automotive sector in the Czech Republic. The high importance of the sector leads to the question of what efficiency local producers achieve.

The evaluation of the efficiency of the automotive sector in the Czech Republic is based on the data envelopment analysis method (DEA). This is a non-parametric method of evaluating efficiency, which is very often used for evaluating the production efficiency of decision-making units (DMU), see for example [4], [5] or [6]. This method has already been applied to the automotive industry several times in the past. However, these studies very often focused mainly on the American or Chinese market. The most recent studies on this topic include [7], [8] or [9]. Among the frequent findings of these studies is the fact that one of the main determinants of efficiency is the size of the enterprise, and that the technical efficiency of producers increases over time.

The main objective of this paper was to evaluate the level and development of the technical efficiency of the automotive industry in the Czech Republic. Regarding the above-mentioned studies, the evaluation of efficiency was carried out with an emphasis on the size of the enterprises. In addition to evaluating the development of efficiency over time, attention was also paid to changes in the production possibility frontier.

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xkovarn1@mendelu.cz.
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, michaela.stankova@mendelu.cz.

2 Material and Methods

The financial data on the enterprises (NACE code 29) were obtained from the Orbis database. In this database, it is possible to obtain accounting data at an annual frequency. The period between 2010 and 2019 was selected for the evaluation. In total, it was possible to include 114 enterprises in the analyses. Employee costs and total company assets were selected as input variables representing labor and capital factors. Both variables are therefore expressed in monetary units (thousands of EUR). According to [10], the use of employee costs can distinguish differently set salaries in individual enterprises and provides better information in the model than in the case of using the number of employees, as in [11]. Total assets in the model represent the physical form of the enterprise's total capital, similarly to [12]. The output variables represent the overall performance of the enterprise in financial units (thousands of EUR) expressed as the size of total sales and in the profit and loss statements of individual years, like in [9].

The radial method of measurement was chosen to calculate (in)efficiency. [8] recommends using input-oriented models for the automotive sector, because according to them, enterprises in this sector are particularly able to influence their inputs. When using aggregated data (such as in [13]), models assuming constant returns to scale are used. Since our dataset contains enterprises of different sizes, so that economies of scale can be distinguished, we decided to use the so-called BCC model (originally developed by [14]), which allows variable returns to scale to be incorporated, as in [7]. The BCC input-oriented model compiled for unit *H*, which is one of the *p* units, is as follows:

$$\begin{aligned}
\max \quad & E_H = \sum_{j=1}^{n} v_{jH} y_{jH} + \mu_H, \\
\text{subject to} \quad & \sum_{i=1}^{m} u_{iH} x_{iH} = 1, \\
& -\sum_{i=1}^{m} u_{iH} x_{ik} + \sum_{j=1}^{n} v_{jH} y_{jk} + \mu_H \leq 0, \forall k = 1,2,\dots,p, \\
& u_{iH} \geq \varepsilon, \forall i = 1,2,\dots,m, \\
& v_{jH} \geq \varepsilon, \forall j = 1,2,\dots,n, \\
& \mu_H \text{ free},
\end{aligned}$$

where the input variable is arranged in matrix $X = \{x_{ik}, i = 1, 2, \dots, m, j = 1, 2, \dots, p\}$, the output variable is arranged in matrix $Y = \{y_{ik}, i = 1, 2, \dots, n, j = 1, 2, \dots, p\}$, ε is the so-called infinitesimal constant, and μ_H represents the magnitude of the deviation from constant returns to scale.

BCC models were constructed for each year separately. To display changes in efficiency over time, the Malmquist index was used, similarly to [15]. This index includes a total change that can be divided into two components, i.e., a change in efficiency and a change in the production possibility frontier. Whether we focus on partial components or the overall Malmquist index, the interpretation of these results is the same as for other indices. Therefore, values greater than 1 indicate an improvement, less than 1 a deterioration, and if the situation did not change, the enterprise would receive a value exactly equal to 1.

To calculate this index, it is necessary to calculate a total of four DEA models (in our case, input-oriented BCC models). All the models and the Malmquist index itself were compiled using DEA SolverPro (version 15f). This software makes it possible to calculate not only the change in two adjacent periods, but also to recalculate the results to the total change (i.e., from the first period to the last). Technical details about this procedure can be found in [16]. The Malmquist index (*MI*) can be defined as the geometric mean of two efficiency ratios (*E*), where one is the efficiency change measures by the period 1 technology and the other is the efficiency change measured by the period 2 technology:

$$MI = \left[\frac{E^1((x_H, y_H)^2)}{E^1((x_H, y_H)^1)} * \frac{E^2((x_H, y_H)^2)}{E^2((x_H, y_H)^1)} \right]^{1/2}.$$

3 Results and Discussion

Based on the results obtained from the BCC input-oriented model, technical efficiency was evaluated according to individual size categories of enterprises. The achieved results of individual categories are shown in Figure 1. This is the median efficiency of individual categories of enterprises in the Czech Republic. In general, it can be stated that the majority of enterprises were identified by the model as being inefficient. It can be seen that the lowest median technical efficiency was achieved by medium-sized enterprises, while the highest technical efficiency was generally achieved by very large and small enterprises. It should be noted that there was a relatively high deviation for small enterprises during the monitored period (discussed in detail below). These strong changes in the achieved efficiency correspond to the finding in [9], where the largest dispersion of technical efficiency was achieved by small enterprises, and on the contrary, large and very large enterprises achieved a relatively low dispersion.

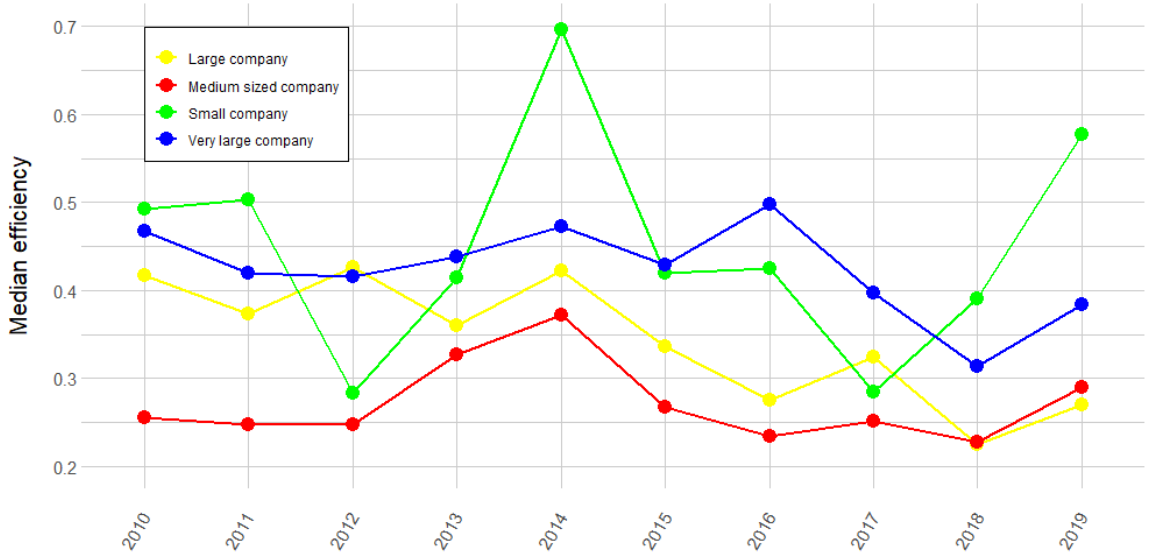


Figure 1 Median efficiency results by enterprise size based on the BCC input-oriented model

The Malmquist index was used to evaluate the overall development of the efficiency of the enterprises. The median results for the overall Malmquist index between 2010 and 2019 and also for individual components of this index between 2010 and 2019 can be found in Table 1. The results in Table 1 show the overall change from the first to the last monitored period. Values greater than 1 indicate improvement during the observed period, while values less than 1 indicate deterioration. It can be seen that, according to the median values, there was a general decrease in individual efficiency for all enterprise sizes, but on the contrary, there was an increase in production possibilities for all enterprise sizes. As a result, in terms of the overall situation (i.e., in terms of the overall index), small, medium-sized, and very large enterprises generally improved their situation. Only for large enterprises, the negative impact of the catch-up effect outweighed the positive impact of the frontier shift, and the overall Malmquist index is less than 1. Since large enterprises make up a large part of the dataset, the Malmquist index for the entire sector is also slightly below the threshold of 1.

Enterprise	Catch-up	Frontier shift	Malmquist index
Small	0.8081	1.4336	1.0668
Medium	0.8068	1.2209	1.0021
Large	0.6761	1.3244	0.9097
Very large	0.7764	1.4857	1.0595
Total	0.7242	1.3679	0.9878

Table 1 Median change in individual efficiency (catch-up), median change in the production frontier (frontier shift), and median change in the Malmquist index between 2010 and 2019

The detailed results in changes for each year are shown in Figure 2. These are the median values for the overall index and its two components (as in Table 1) in the year-on-year variant.

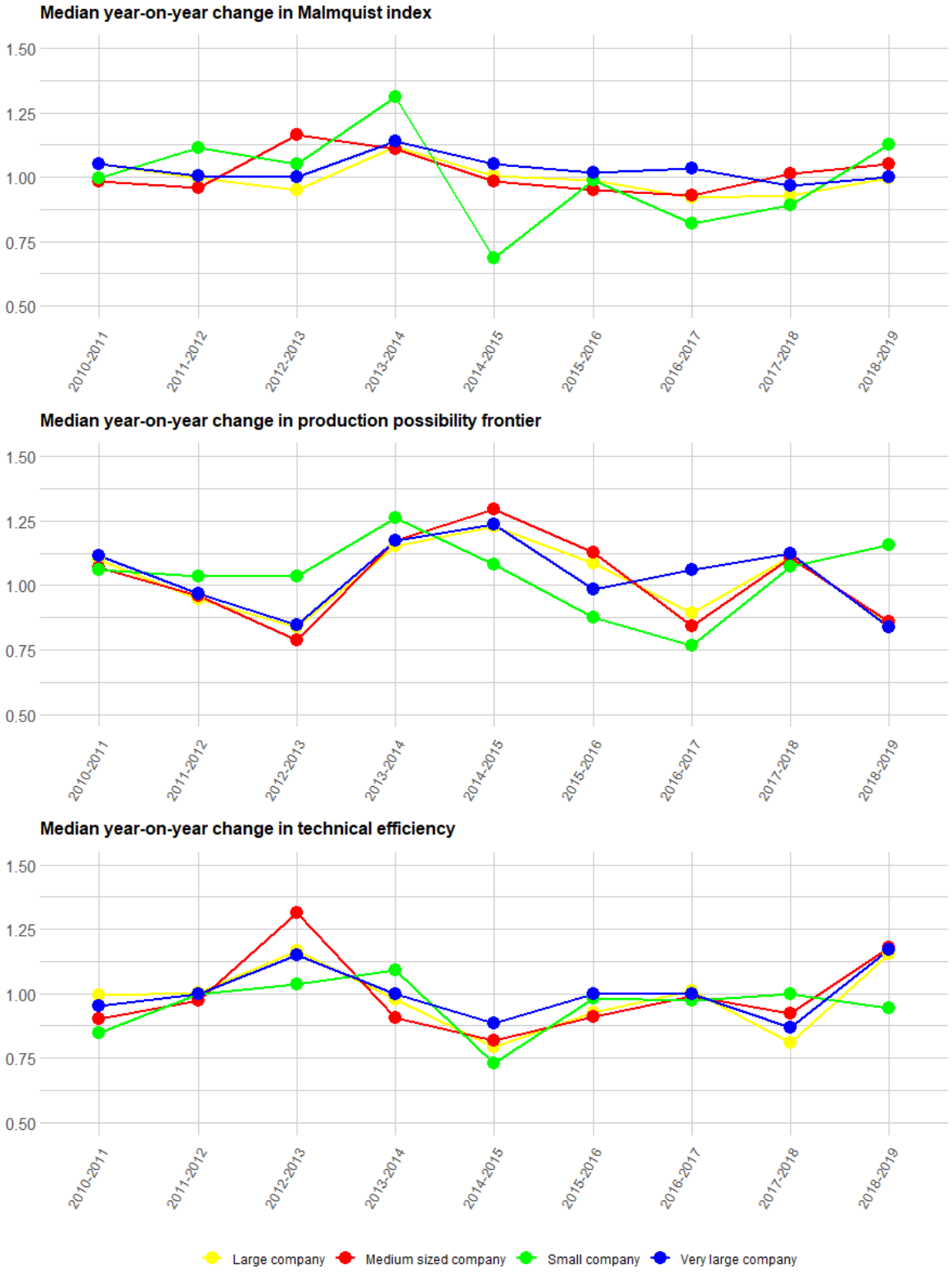


Figure 2 Development of changes in the Malmquist index technical efficiency, production possibility frontier, and technical efficiency

In general, we see the largest differences in the development shown in Figure 2 for small enterprises. According to [11], small manufacturing enterprises have an advantage in their flexibility, whereby they can respond more quickly to market changes. The different development, especially in their production possibilities, and

consequently also within the entire Malmquist index, for small enterprises in the automotive sector can be explained precisely by their different flexibility.

After the financial crisis in 2009, other (larger) enterprises show a systematic decrease in the production possibility frontier until 2012. According to [15], 2012 was a turning point for enterprises, as for the first time after the financial crisis there was an increase in GDP in the Czech Republic. From this point on, the economy began to demonstrably recover and enterprises in all economic sectors could start to prosper, which was positively reflected in the growth of their production possibilities, as our results for the automotive sector in Figure 2 also demonstrate. Small enterprises, which typically focus on very specialized activities/products, have relatively secured sales even in bad times. At the same time, they are able to save a large amount of their costs even when dismissing just a single employee. Thanks to this, they are able to adapt better and faster to the market situation compared to their larger competitors. Given that these are typically family businesses that often subsidize the business from their own family's financial reserves, small- enterprises are more optimistic about making investments (for example, buying new technologies) even in times of crisis. All these specific characteristics together caused small enterprises to have a different development in Figure 2. However, as soon as the general economy in the Czech Republic began to prosper, even larger enterprises could afford to invest part of their profits in investments and therefore increase their production possibilities after 2012. In addition, according to the results in Table 1, very large enterprises generally contributed the most to the growth of production possibilities between 2010 and 2019.

The development of median individual efficiency is then a reflection of the situation with the development of changes in production possibilities. In a period when, due to the effects of the financial crisis, the production possibility frontier fell, enterprises paradoxically got closer to the efficiency frontier. On the contrary, in times when enterprises could implement new technologies and the production possibility frontier increased, the efficiency of other enterprises decreased because they moved further away from this shifted frontier.

4 Conclusion

In this article, an analysis of the automotive industry in the Czech Republic was conducted. Individual enterprises were divided into four categories according to their size. Subsequently, the development of changes in technical efficiency, the production frontier, and the Malmquist index was monitored in the period between 2010 and 2019. Our results show that dividing enterprises according to their size makes sense, as their efficiency (as well as changes in efficiency and changes in production possibilities) varies depending on the size of the enterprise.

The results show that small enterprises differentiate themselves from their larger competitors. In the case of the results of the median technical efficiency, a certain fluctuation in their results can be seen, since in some years small enterprises are among the worst together with medium-sized enterprises, but in other years they are on the contrary in first place. In general, it can also be stated that small, medium-sized, and large enterprises improved their position between 2010 and 2019, mainly thanks to the increase in production possibilities. A subject of future research will be the development of efficiency in the automotive sector during the COVID-19 pandemic. Considering our results, impacts should be evaluated with respect to the size of the enterprise.

Acknowledgements

This article was supported by grant No. IGA-PEF-TP-23-005 of the IGA PEF MENDELU Grant Agency.

References

[1] Zorpas, A. A & Inglezakis, V. J. (2012). Automotive industry challenges in meeting EU 2015 environmental standard. *Technology in Society, 34(1),* 55–83.

[2] Kovárník, R. & Staňková, M. (2021). Determinants of Electric Car Sales in Europe. *LOGI – Scientific Journal on Transport and Logistics, 12(1),* 214–225.

[3] Sdružení automobilového průmyslu. *Obecné základní přehledy o českém automobilovém průmyslu*. [Online]. Available at: https://autosap.cz/zakladni-prehledy-automotive/obecne-zakladni-prehledy/ [cited 2023-03-11].

[4] Staňková, M. & Hampel, D. (2019). Bankruptcy Prediction Based on Data Envelopment Analysis. *Mathematical Methods in Economics 2019: Conference Proceedings. České Budějovice: Jihočeská univerzita v Českých Budějovicích,* 31–36.

[5] Křetínská, M. & Staňková, M. (2021). Evaluation of the Construction Sector: a Data Envelopment Analysis Approach. *Mathematical Methods in Economics 2021: Conference Proceedings. Praha: Česká zemědělská univerzita v Praze*, 287–292.
[6] Mašková, K. & Blašková, V. (2021). Efficiency of tertiary education in EU countries. *Mathematical Methods in Economics 2021: Conference Proceedings*. *Praha: Česká zemědělská univerzita v Praze*, 312–316.
[7] Stefanoni, S. & Voltes-Dorta, A. (2021). Technical efficiency of car manufacturers under environmental and sustainability pressures: A Data Envelopment Analysis approach. *Journal of Cleaner Production, 311,* 127589.
[8] Jiang, H., Han, L., Ding, Y. & He, Y. (2018). Operating efficiency evaluation of China listed automotive firms: 2012–2016. *Sustainability, 10(1),* 184.
[9] Kovárník, R. & Staňková, M. (2023). Efficiency of the Automotive Industry in the Visegrad Group. *LOGI – Scientific Journal on Transport and Logistics, 1(1),* 12–23.
[10] Staňková, M. & Hampel, D. (2018). Efficiency Comparison in the Development of Building Projects Sector. *Mathematical Methods in Economics 2018: Conference Proceedings. Praha: MatfyzPress*, 503–508.
[11] Staňková, M., Hampel, D & Janová, J. (2022). Micro-data efficiency evaluation of forest companies: The case of Central Europe. *Croatian Journal of Forest Engineering, 43(2)*, 441–456.
[12] Staňková, M. & Hampel, D. (2020). Efficiency Assessment of the UK Travel Agency Companies – Data Envelopment Analysis Approach. *Mathematical Methods in Economics 2020: Conference Proceedings. Brno: Mendelova univerzita v Brně,* 550–556.
[13] Kabát, L., Hampel, D., Grochova, L. I., Janová, J. & Střelec, L. (2014). Alternative approaches for assessing the European countries economic and social results. *International Conference on Enterprise and the Competitive Environment (ECE). Brno: Procedia Economics and Finance, 12,* 273–282.
[14] 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.
[15] Krejčí, M & Staňková, M. (2022). The Position of the Czech Republic within the Metallurgical Sector. *Mathematical Methods in Economics 2022: Conference Proceedings*. *Jihlava: Vysoká škola polytechnická Jihlava*, 193–198.
[16] Cooper, W. W., Seiford, L.M. & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA/Solver Software*. New York: Springer Science & Business Media.

Analysis of the Demand for Local Food in the Czech Republic by Applying the Theory of Planned Behavior

Petra Králová[1], Jana Krajčová[2]

Abstract. This paper aims to explore the motivations to buy locally farmed food. We used the tools from the Theory of Planned Behavior to identify the inherent factors that drive consumer choices. We focused on Gen Z, as this can provide important insights on a specific and important customer segment. The motivations for buying locally farmed food vary, in general and across generations. Some believe that locally farmed food is healthier because it contains fewer chemicals, others stress that it is closely linked to the natural environment and its biochemical composition. Some consumers are socially responsible and they want to support their community and the local farmers or reduce their carbon footprint. For younger generations, buying behavior is also part of the lifestyle, their personality and uniqueness. We conducted a survey of 232 Generation Z young adults above the age of 17, to inquire about their motivations and preferences to buy locally farmed food. Our questionnaire and the general methodology based on theory of planned behavior extended with the variable "uniqueness-seeking lifestyle" follows the approach of Ham [18], but focuses on locally farmed food and the Czech market in particular. We conducted the exploratory factor analysis, in order to identify which responses correlate the most with respondents' latent motivations to buy locally farmed food. Preliminary findings confirm intuitive expectations and help us link responses to individual questions to latent constructs of behavioral beliefs, attitudes, subjective norms, perceived behavioral control, uniqueness seeking lifestyle and commitment. Understanding the motivations for buying locally farmed food can help us design more successful marketing strategies which will not only help to support local farmers but also the local environment and the overall sustainability.

Keywords: locally farmed food, theory of planned behavior, Gen Z

JEL Classification: C44
AMS Classification: 90C15

1 Introduction and Motivation

In this paper, we study the motivations for purchasing locally farmed foods of Generation Z (people born in 1996 and later) in the Czech Republic.

Healthy nutrition, proper diet and sustainability of food production and consumption are now increasingly emphasized topics. Customers choose very carefully the food they buy for themselves and their families. They are interested in the composition of food, its energy value, the origin of food and the way it is grown. There are generational differences, though. The consumer choices in the Czech Republic have expanded dramatically after the fall of communism and the youngest consumers now do not know the lack of quantity and of quality that their parents and grandparents used to automatically accept and deal with. With growing awareness of the impact of human activity on global environment and human health and tremendous improvement in access to information, the Czech consumers are now choosing much more carefully and responsibly. This is, however, still more true for the young ones than for the older ones. Understanding the beliefs and intrinsic motivations of young consumers is therefore crucial not only for successful marketing but also for supporting their environmental, social and individual responsibility.

Local food is food that is distributed from the grower or farmer to the customer within a relatively small geographic area. It does not travel long distances, typically does not leave the country of origin. It is therefore fresh and with

[1] Vysoká škola chemicko-technologická v Praze, Ústav ekonomiky a managementu, petra.kralova@vscht.cz.
[2] Vysoká škola chemicko-technologická v Praze, Ústav ekonomiky a managementu, jana.krajcova@vscht.cz

smaller transportation footprint. Big part of local food is also organically farmed which makes it even more attractive for presumed better quality and even lower overall environmental footprint. In the Czech Republic, the locally farmed food is typically sold in local open-air markets, specialized sections of some supermarket chains, small farm shops or online through e-shops. In this paper, we specifically look at basic items - milk and dairy products, eggs, meat (pork), fruits, vegetables (apples, potatoes), baked goods, flour, pulses and alcoholic beverages.

The market for locally farmed food is relatively new but it has grown in importance in recent years; the variability of choice and ease of access increased visibly. From statistical data in the Czech Republic [27], but also worldwide, it is obvious that there is a growing interest in organic food and a growing interest in buying food from local sources. Total consumption of organic food (including imports) in the Czech Republic was CZK 6.15 billion (EUR 0,26 billion), of which local food sold through farms amounted to CZK 330 million (EUR 13,75 million) and (Local and other organic food sold through e-shops CZK 1,293 million (EUR 0,054 million) in 2021.

The arguments for buying local food are many. If the food originates from local environment, the customer knows about the conditions in which it has been grown. It is also biologically natural to eat food that has been grown in the region for centuries and the human organism is closer to it from evolutionary point of view. Traditions of eating and preparing food are based on locally available food. Shopping locally helps to support local growers/farmers and thereby the entire region economically. Local food is not transported over long distances and therefore it is fresh and does not need special preservatives. Less transportation means lower CO2 emissions. But the overall environmental footprint of the local farm food market is, in general, lower. Last, but not least,2e all of the above is not only important from the sustainability perspective, but for many people it becomes a lifestyle, a sense of belonging to community of responsible humans, a way of feeling good about making the right choices and living in certain harmony with nature.

We hypothesize that all of the motivations are best combined in the young generation which has benefited from the fast development of information technologies and has been exposed all their lives to news about devastating impacts of human activity (and especially modern agriculture) on global environment. Their involvement in social networks puts special pressure on them; they seek the ideas to identify themselves with, they often fight for uniqueness. We argue that over the recent decades, both the food market and the customers – their needs, moral values and preferences – have changed and developed substantially, which is mostly true for the young ones. Therefore, our paper focuses on the Generation Z customer segment. It is the interest of young people (aged up to 27) that suggests the importance of locally farmed food and its potential for this group and beyond. Gen Z is at an age where they are setting or adjusting their diet and lifestyle not only for themselves but also for their future families or for their young children and of course this generation has a potential to change old habits of their parents and grandparents. Thus, there is considerable economic potential for locally farmed foods if their popularity among the Gen Z population grows.

Gen Z

Labelling generations with letters is used in demography, social sciences, marketing or popular culture. Typically, particular generation is closely linked to certain historical events and consequent social or technological changes which likely played a big role during their upbringing. Looking at an issue through generations makes it possible to analyze changes in behavior in the long term. Distinguishing generations can provide a way to understand how different formative experiences interact with the life-cycle and aging process to shape people's views of the world. Generations are often considered by their span, but again there is no agreed upon formula for how long that span should be. But for analytical purposes, we adopted 1996 as a meaningful cutoff between Millennials and Gen Z for a number of reasons, including key political, economic and social factors that define the Millennial generation's formative years. According to Dimock [13], Generation Z is a generation whose behavior and even life-choices have been affected by the use of technology from early childhood (iPhone was launched in 2007) and by the use of social networks. Moreover, Generation Z will likely show some strong consumer-oriented differences from Generation Y because of the age of these individuals during periods of economic recession.

Wood [32] offers four trends that are likely to characterize Generation Z as consumers:

1. a focus on innovation;

2. an insistence on convenience;

3. an underlying desire for security;

4. a tendency toward escapism.

Wood [32] also suggests that Generation Z are the children of Generation X, who in many cases were subjected to parental divorce and had to fend for themselves, there was pressure to be independent from young age. This is the reason why Generation X tends to focus on products that make life easier, such as convenience products or why they tend towards simplistic solutions. According to the research by Freedonia Group [16], Gen Z and millennials, are generations, who are 29% more likely to try new products than other age cohorts. They also tend to be the biggest users of natural and organic food. Moreover, adults under the age of 40 are especially prone to value healthy, premium, and fresh foods, including organic produce and meat. Younger consumers including Gen Z, Millennials, and younger members of Generation X are the most likely to say they tend to buy organic produce to varying degrees. As Gen Z in particular obtains a higher income, often after graduating from college, they are expected to partake of more organic produce and become more significant buyers [16].

2 Methods and Data

2.1 Theory of Planned Behavior

In this paper, we focus on explaining the motives of local food consumption using the theory of planned behavior. The theory of planned behavior is a psychological/behavioral theory that links beliefs to behavior and can therefore be helpful for understanding various motivations behind the consumer choices. Social attitudes, personality traits, and cognitive self-regulation especially, have played an important role in the attempts to predict and explain the human behavior [2, 3, 10, 22]. The theory of planned behavior is also designed to predict and explain human behavior in specific contexts [3]. It originated as an extension of the theory of reasoned action [4, 15] necessary due to the original model's limitations in dealing with behaviors over which people have incomplete volitional control.

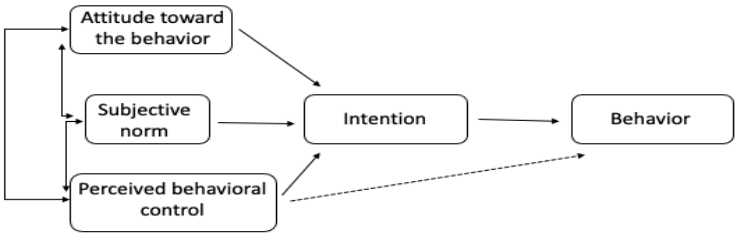


Figure 1 Theory of planned behavior (Ajzen 1991)

The core concept of the theory are intentions. Intentions are assumed to capture the motivational factors that influence a behavior. They are indications of how hard people are willing to try, of how much of an effort they are planning to exert, in order to perform the behavior [3]. *Attitudes* toward the behavior, *subjective norms* with respect to the behavior, and *perceived control* over the behavior are the three components that together shape individual behavioral intentions. Empirically, they are usually found to predict behavioral intentions with a high degree of accuracy and together can account for a considerable proportion of variance in behavior [3]. The trend in the literature is to view the sustainable consumption and purchase of organic food as a way for consumers to comply with the norms of society [8, 21, 28], which provides a support for connecting the topic of locally farmed food with methods of theory of planned behavior.

The theory of planned behavior has attracted many critics. These investigators tend to deny the importance of consciousness as a causal agent [29, 30] and view much human social behavior as driven by implicit attitudes [17] and other unconscious mental processes [1, 6, 7, 9, 26]. The evidence, however, suggests that the theory of planned behavior does in fact predict intentions and behavior quite well. Investigators have recently turned their attention to more sophisticated questions, although straight-forward applications to new behaviors or behaviors in novel settings continue to appear in print [5].

In this paper, we examine the direction and the strength of the effects of inherent factors behind the intention to purchase local food and behind the referent actual behavior. In line with the theory of planned behavior, the intention to buy is directly influenced by three factors, which are also known as the antecedents of intention: *personal attitudes*, *subjective norms* and *perceived behavioral control*. *Personal attitudes* are based on the beliefs that an individual possesses about a specific object coupled with attributes attached to that object. Various studies on the relationship between attitude and the intention to buy have consistently shown positive associations (for example,

[11, 25, or 12]) reported that a positive attitude supports a consumer's intention to purchase organic food. We expect similar to be true for locally farmed food. *Subjective norms* represent one's beliefs about whether the people of their social circle would encourage, approve, disapprove specific behavior. Analogically, we believe that respondents whose peers approve or, in fact, routinely purchase locally farmed food, will also exhibit similar preferences or behavior. *Perceived behavioral control* refers to one's perception of how easy or how difficult it is to engage in the behavior of interest. Easy and plentiful access to locally farmed food is likely to positively affect the intentions to buy locally farmed food.

Following the methodology of [18] we expanded the core theory to explore additional behavioral concepts such as uniqueness-seeking lifestyle as an additional motivational factor. We conducted a questionnaire survey, in which the respondents were inquired not only about their shopping habits, but also about the motivations and norms of their own and of their social circle. All the questions are on 1-5 Likert scale, where 1 corresponds to "strongly agree" and 5 to "strongly disagree."

In the next step, we used exploratory factor analysis, which is a useful and appropriate statistical method to identify the latent factors (behind the planned behavior) based on the responses in the questionnaire.

2.2 Exploratory Factor Analysis

Factor analysis is a multivariate statistical method, which is used to analyze the structure of interdependencies of variables. The analysis assumes that these dependencies are the result of a small number of background unmeasurable factors, which are called common factors (or latent variables). Factor analysis thus explains the linear interdependence of observed variables by the existence of a small number of unobservable factors and other sources of variability called error components. In factor analysis, the researcher works with set of measured variables (such as responses in questionnaires) and aims to identify the items with the strongest relationship to the latent and to reveal the structure of this relationships. As a result, the number of variables can be reduced, and the identified structure can be used in further research to model and predict the latent construct variable.

Problems in factor analysis may lie in the ambiguity of factor estimates (for example due to the dependence of results on the rotation used, or on particular sample) or in the final number of identified latent factors (based on the mixture of statistical results and the researcher's expertise).

There are two main approaches: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is essential to establish underlying constructs for a set of measured variables. It is typically used, when there is no clear guidance provided by the theory and the researcher needs to rely on collected data and the statistical results obtained from it to identify the underlying structure. CFA allows the researcher with a clearly defined hypothesis about the relationship between the observed variables and their underlying latent constructs to test this relationship and its statistical significance (e.g., [19 or 20]).

In formulating the underlying models of theory of planned behavior as a framework for understanding the Czech Generation Z's demand for local food, we adopted as a starting point the methodology of Ham [18]. Since the research questionnaire contained questions which could have been intuitively linked to several different latent constructs, in this paper we employed exploratory factor analysis in order to discover and identify the underlying structures. Responses to all 30 survey questions were grouped intuitively to construct measures of behavioral beliefs, attitudes, subjective norms, perceived behavioral control, uniqueness seeking lifestyle, and commitment.

For each latent construct we first tested for suitability of used approach using Bartlett's test of sphericity (BTS), Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) and Cronbach's alpha coefficient (CAC). The null hypothesis of the Bartlett's test is that the variables are not correlated (and thus the data would not be suitable). The reference values of KMO measure are as follows: <0.49 suggest unacceptable sampling adequacy, 0.50-0.59 miserable, 0.60 to 0.69 mediocre, 0.70 to 0.79 middling, 0.80 to 0.89 meritorious and above 0.90 a marvelous sample adequacy. Cronbach's alpha is a reliability coefficient, attaining values between 0 and 1; the values above 0.7 are typically considered high enough for early stages of research (help for Stata software [24]).

Consequently, we identified for each of the presumed latent constructs the groups of survey questions which could be intuitively related to that construct and run the exploratory factor analysis. The relevant statistical results are presented in the following section, where each of the latent constructs is discussed individually. Each of the questions was then assigned to factor for which it showed the highest loading, questions with loadings below 0.3 were considered as not relevant for that factor. Only factors with non-negative eigenvalues have been retained (eigenvalues are not reported here as they are relevant for factor retainment but not for practical interpretations). The analysis was conducted using the Stata software [23].

3 Results with Interpretations

We conducted our analysis on sample of 232 generation Z young adults of age above 17. Average age in our sample was 20.71 with 150 female, 80 male and 3 respondents who wished not to specify their gender. The survey was conducted in the Czech Republic, however our sample also contains foreign students. About 20 percent of our respondents have spent less than one year in the Czech Republic prior to participating in our survey. As this may interfere with their responses to some of the survey questions, this fact will be accounted for in further analysis. We realize that some of our respondents are young enough to be still living with their parents and therefore we also collected information on the their household (and shopping) arrangement. The results are summarized in Table 1 below. The most critical group is the first one, which still lives with their parents. However only 29 of these 68 participants responded in the subsequent question that their parents or a partner are shopping groceries for them. Since the responses of these individuals to some of the questions in our survey can be only considered hypothetical, this subgroup will be either discarded or treated separately in further analysis.

Type of living	**Freq.**	**Percent**	**Cum.**
still with parents	68	30.63	30.63
dormitory	73	32.88	63.51
rented with non-relatives	49	22.07	85.59
rented alone	15	6.76	92.34
own/family owned flat	15	6.76	99.1
other	2	0.9	100

Table 1 Which of the following best describes your current living situation?

The results of pre-tests, the p-value for Bartlett's test of sphericity (BTS) and Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO), as well as number of retained factors, factor loadings and Cronbach's alpha (CAC) coefficients for each construct together with brief interpretations are provided below.

Construct 1: Behavioral Beliefs

The statistical results and the questions that represent relevant measured responses can be found in Table 2 below. Number of retained factors is two. Questions 2 to 4 load the most on factor 1, which can be interpreted as *responsibility towards society* and question 1 loads most on factor 2 which can be interpreted intuitively as *responsibility towards oneself*. Question 2 also loads relatively highly on the second factor. The uniqueness value, which represents the percentage of the variance that is 'unique' to the variable and not shared with other variables, is the highest for question 1. The pretests suggest that sample is suitable for factor analysis (the KMO measure of sampling adequacy suggests only middling sufficiency). The signs and magnitudes of loadings suggest a positive correlation between measured responses to questions targeting individual and social responsibility and unmeasurable behavioral beliefs.

	BTS	**0.0000**	KMO	**0.753**	CAC	**0.7787**
Loadings	factor 1	factor 2				Uniqueness
Q1. *Buying locally farmed food enables me to contribute to my health and the health of my family.*	0.1663	**0.4097**				0.7018
Q2. *Buying locally farmed food enables me to help protect the environment.*	**0.4198**	0.3612				0.4647
Q3. *Buying locally farmed food enables me to be socially responsible.*	**0.6359**	0.1276				0.4571
Q4. *Buying locally farmed food enables me to soothe my conscience regarding the damage that we are doing to future generations.*	**0.6345**	0.0868				0.5069

Table 2 Behavioral beliefs

Construct 2: Personal Attitudes

The statistical results and the included questions are in Table 3 below. Three factors fulfilled the criteria for retainment. Questions 5 to 7 load the most on factor 2, which can be interpreted as *focus on quality of products*.

Questions 8 to 10 load most on factor 1 which can be interpreted as *inherent preference for local food*. Questions 24 and 25 load the most on factor 3, which can be interpreted as *belief that benefit exceeds the cost for local food*. Uniqueness values are the lowest for questions 5, 6 and 10, suggesting the highest relevance of these items in the factor model. The pretests suggest that sample is suitable for factor analysis (the KMO measure suggests only middling sufficiency). The loadings of questions are bolded in the Table 3 to indicate their assignments to factors; they all show positive correlations with the unmeasurables. This is to be interpreted as positive effect of personal attitudes on intentions to buy local food, where personal attitudes are represented by measured responses regarding the questions that target *focus on quality* of the products, *inherent preference* for local food and the *belief that benefit exceeds the cost*.

	BTS	**0.000**	KMO	**0.728**	CAC	**0.7596**
Loadings	factor 1	factor 2	factor 3			Unique-ness
Q5. *When purchasing food, I check its quality certificates.*	-0.046	**0.7895**	0.0109			0.4384
Q6. *When purchasing food, I check its origin certificates.*	-0.0196	**0.7582**	0.0481			0.3349
Q7. *When purchasing food, I check its ingredients.*	0.1677	**0.3603**	-0.0629			0.658
Q8. *Buying locally farmed food would give me great satisfaction.*	**0.3457**	0.1561	0.3037			0.5874
Q9. *If I had the opportunity and necessary resources, I would purchase locally farmed food.*	**0.7097**	-0.0404	-0.0281			0.5113
Q10. *Among different nutrition choices, I would rather choose the locally farmed food.*	**0.6648**	-0.0339	0.1261			0.4433
Q24. *I am willing to buy locally farmed food because the benefits outweigh the cost.*	-0.0499	0.0505	**0.566**			0.692
Q25. *Buying locally farmed food is the right choice even though they cost more.*	0.1514	-0.0786	**0.4812**			0.6833

Table 3 Personal attitudes

Construct 3: Subjective Norms

The statistical results and relevant questions are in Table 4. Only one factor has been retained, with reasonably high loadings on all questions, which is intuitively in line with assumed latent variable *subjective norms* and their positive correlation with measured responses. Uniqueness values for all items are medium sized. The pretests suggest that sample is suitable for factor analysis, even though the KMO measure of sampling adequacy suggests only mediocre sufficiency. Cronbach's alpha coefficient above 0.7 is typically considered high enough for early stages of research.

	BTS	**0.000**	KMO	**0.694**	CAC	**0.7707**
Loadings	factor 1					Uniqueness
Q11. *My friends think I should buy locally farmed food.*	**0.6473**					0.5810
Q12. *My relatives think I should buy locally farmed food.*	**0.7217**					0.4792
Q13. *My family thinks I should buy locally farmed food.*	**0.7035**					0.5051

Table 4 Subjective norms

Construct 4: Perceived Behavioral Control

The statistical results and tested questions are in Table 5. Questions 14 to 17 load high on factor 1, while the loading of question 15 indicates a negative correlation with the factor. This is intuitively appropriate given the wording of the question. Question 17 loads reasonably on both factors 1 and 2, with greater loading on factor 1. Here, we will however follow the intuition over statistical results and thus assign questions 14 to 17 to intuitive construct *affordability of locally farmed food*, and questions 18 and 19 to construct *local availability of LFF*. Questions 15, 16 and 18 have the lowest relevance in the factor model according to uniqueness values. The pretests suggest that sample is suitable for factor analysis, however with mediocre sufficiency of the KMO measure and borderline Cronbach's alpha. Conducting further analysis on bigger and more variable sample is therefore necessary to obtain more reliable results. All-in-all affordability, positive economic situation and physical availability contribute to intentions to buy local food through perceived behavioral control.

	BTS	**0.0000**	KMO	**0.665**	CAC	**0.6575**
Loadings	factor 1	factor 2				Unique-ness
Q14. *I have enough money to buy locally farmed food.*	**0.6912**	0.0317				0.4532

Q15. *Current economic situation negatively affects my ability to buy locally farmed food.*	**-0.3893**	0.0307		0.7226
Q16. *I would still buy locally farmed food, even if the economic situation was to worsen.*	**0.6110**	-0.0375		0.6571
Q17. *I have the ability to buy locally farmed food, they are available.*	**0.5152**	**0.3113**		0.5254
Q18. *I know where to buy locally farmed food.*	0.0502	**0.4259**		0.8078

Table 5 Perceived behavioral control

Construct 5: Uniqueness Seeking

The statistical results are in Table 6. Survey questions 19 to 21 have been included into preliminary testing. Pretests suggest lower sample adequacy, KMO of 0.573 is in general not sufficient for reasonable analysis. The computations below thus constitute mere preliminary findings that are yet to be confirmed by further analysis. Cronbach's alpha between 0.6 and 0.7 is also acceptable only for primary stages of research. Only factor 1 passes the criteria for retainment, with highest loadings on questions 19 and 20 which have direct and explicit interpretation as *uniqueness seeking*. Uniqueness value is also quite high and loading low for question 21, suggesting its low relevance in the factor model. Based on the results, the uniqueness seeking plays a role in motivation to buy local food.

	BTS	**0.000**	KMO	**0.573**	CAC	**0.6932**
Loadings	factor 1					Uniqueness
Q19. *Buying locally farmed good enables me to be different and to emphasize my different lifestyle.*	**0.6761**					0.5625
Q20. *Buying locally farmed food is an important part of my personality.*	**0.6618**					0.42
Q21. *I would still buy locally farmed food even if conventional alternatives were on sale.*	0.2692					0.7706

Table 6 Uniqueness seeking

Construct 6: Commitment

The statistical results and included questions are in Table 7. Survey questions 21 to 25 have been considered in the preliminary analysis of commitment to buy locally farmed food. Two factors have been retained. Questions 21 and 22 load highest on factor 2, which intuitively correspond to *strong inherent commitment for LFF*. Questions 23 to 25 correspond to *commitment based on personal attitudes and norms*. Results of preliminary tests suggest reasonable sample adequacy and suitability for factor analysis. Relatively high uniqueness values suggest weaker role of individual items in factor model. The factor loadings suggest a medium positive correlation of measured responses with the unmeasurables.

	BTS	**0.0000**	KMO	**0.740**	CAC	**0.7188**
Loadings	factor 1	factor 2				Uniqueness
Q21. *I would still buy locally farmed food even if conventional alternatives were on sale.*	0.1102	**0.5821**				0.5764
Q22. *I am willing to do whatever it takes to buy locally farmed food.*	0.069	**0.5975**				0.5915
Q23. *I have seriously considered to start buying more locally farmed food products.*	**0.5004**	0.2019				0.5946
Q24. *I am willing to buy locally farmed food because the benefits outweigh the cost.*	**0.5713**	0.193				0.5116
Q25. *Buying locally farmed food is the right choice even though they cost more.*	**0.5252**	-0.0998				0.7736

Table 7 Commitment

4 Conclusion

Preliminary analysis of survey data intended to identify latent variables affecting the motivations to buy locally farmed food was conducted using explanatory factor analysis. Preliminary findings confirm intuitive expectations and help us link responses to individual questions to latent constructs of behavioral beliefs, attitudes, subjective

norms, perceived behavioral control, uniqueness seeking lifestyle and commitment. We identified factors of responsibility towards oneself and towards society, which can explain positive correlation between the questions of behavioral beliefs category. For the role of personal attitudes, two factors, focus on quality and inherent preference for local food explain positive correlation between the relevant group of questions. Questions from the area of subjective norms can be represented by one factor, which explains reasonable part of correlation between them. Uniqueness seeking can also be represented by one factor, with reasonable relevancy. The questions from the area of perceived behavioral control resulted in two factors, affordability and physical availability of the locally farmed food, the results suggest that current economic situation has a non-negligible effect on ability to purchase locally farmed food. The results on commitment suggest that part of motivation stems from non-specific inherent commitment, but another part is closely related to personal attitudes and norms.

For some variables the results are statistically stronger than for others and therefore further analysis on extended sample will be conducted. The next step will be confirmatory factor analysis with structural equation modelling approach to estimate the latent variables with grouped questions (as suggested by the results of EFA). The estimated latent variables will be then used to test hypotheses in order to get further insight on the roles of individual motivations to purchase locally farmed food. The collected background characteristics will be included in the analysis in order to understand the role of additional factors such as gender, education, or size of town. The results can be then used to design more efficient marketing strategies or policy measures to support local farms (not exclusively) in the Czech Republic.

References

[1] Aarts, H. & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal directed behavior. *Journal of Personality and Social Psychology*, 78, 53–63.

[2] Ajzen, I. (1988). *Attitudes, personality, and behavior.* Chicago: Dorsey Press.

[3] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211.

[4] Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.

[5] Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & health, 26(9)*, 1113-1127.

[6] Bargh, J.A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J.S. Uleman & J.A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York, NY: Guilford.

[7] Bargh, J.A. & Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462–479.

[8] Biel, A. & Thøgersen, J. (2007), Activation of social norms in social dilemmas: a review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, Vol. 28 No. 1, pp. 93-112.

[9] Brandstätter, V., Lengfelder, A. & Gollwitzer, P.M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology*, 81, 946–960.

[10] Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science.* (Vol. 6, pp. 94-172). New York: Mc-Graw-Hill.

[11] Chen, M.F. (2009). Attitude toward organic foods among Taiwanese as related to health consciousness, environmental attitudes, and the mediating effects of a healthy lifestyle. *British Food Journal*, Vol. 111 No. 2, pp. 165-178.

[12] Chih-Ching, T. & Yu-Mei, W. (2015). Decisional factors driving organic food consumption. *British Food Journal*, Vol. 117 No. 3, pp. 1066-1081.

[13] Dimock, M. (2019). *Defining generations: Where Millennials end and Generation Z begins*. Pew Research Center, 17(1), 1-7.

[14] FIBL-AMI SURVEYS. *Data on organic agriculture in Europe*. [Online]. Available at: https://statistics.fibl.org/europe.html [cited 2023-05-12].

[15] Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

[16] Freedonia Group. *Eating Trends: Generational Food Shopping*. [Online]. Available at: https://www.freedoniagroup.com/packaged-facts/eating-trends-generational-food-shopping [cited 2023-05-12].

[17] Greenwald, A.G. & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.

[18] Ham, M., Pap, A. & Stanic, M. (2018). What drives organic food purchasing?–evidence from Croatia. *British Food Journal.* pp. 1-14.
[19] Hebák, P., Hustopecký, J., Pecáková, I., Průša, M., Řezanková, H., Svobodová, A. & Vlach, P. (2007). *Vícerozměrné statistické metody (3)* 2. přepracované vydání, Informatorium, Praha, ISBN 9788073330019.
[20] Manly, B.F.J. (1994). *Multivariate Statistical Methods*. Second edition. Chapman & Hall. 232 pp.
[21] Rivis, A. & Sheeran, P. (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: a meta-analysis. *Current Psychology*, Vol. 22 No. 3, pp. 218-233.
[22] Sherman, S. J. & Fazio, R. H. (1983). Parallels between attitudes and traits as predictors of behavior. *Journal of Personality*, 51, 308-345.
[23] StataCorp. (2015a). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
[24] StataCorp. (2015b). Stata 14 Base Reference Manual. College Station, TX: Stata Press.
[25] Thøgersen, J. (2009). Consumer decision making with regard to organic food products. in Vaz,M.T.D.N., Vaz, P., Nijkamp, P. and Rastoin, J.L. (Eds), *Traditional Food Production Facing Sustainability: A European Challenge*, Ashgate, Farnham, pp. 173-194.
[26] Uhlmann, E. & Swanson, J. (2004). Exposure to violent video games increases automatic aggressiveness. *Journal of Adolescence*, 27, 41–52.
[27] ÚZEI. *Statistická šetření ekologického zemědělství Zpráva o trhu s biopotravinami v ČR v roce 2021* [Online]. Available at: https://eagri.cz/public/web/file/721713/Zprava_o_trhu_s_biopotravinami_v_CR_v_roce_2021.pdf [cited 2023-05-12].
[28] Vermeir, I. & Verbeke, W. (2006). Sustainable food consumption: exploring the consumer attitude behavioural intention gap. *Journal of Agricultural and Environmental Ethics*, Vol. 19 No. 2,
[29] Wegner, D.M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
[30] Wegner, D.M. & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480–492.
[31] Willer, Helga, Trávniček, Jan, Meier, Claudia, Schlatter & Berndhard, (2023). *The World of Organic Agriculture. Statistics and Emerging Trends 2023*. Research Institute of Organic Agriculture (FiBL) and IFOAM – Organic International.
[32] Wood, S. (2013). *Generation Z as consumers: trends and innovation.* Institute for Emerging Issues: NC State University, 119(9), 7767-7779.

Econometric Aspects of Elasticity of Substitution

Petr Krautwurm[1], Michal Černý[2]

Abstract. This paper reviews the current knowledge on the concept of elasticity of substitution in economic analysis and explores its econometric applications. The elasticity of substitution is a vital tool in microeconomics for determining the substitutability of products. Unlike cross-price elasticities, which require adjustments for endogeneity, the elasticity of substitution can be analyzed directly using market optimum data. Additionally, the elasticity of substitution provides information on the degree of substitutability. However, the use of elasticity of substitution is not without limitations, such as the occurrence of perfect substitutes in data or nonlinear pricing. This paper offers solutions for these issues and outlines the key assumptions necessary for correct usage of the estimator for elasticity of substitution.

Keywords: Elasticity of substitution, Estimator, Market optimum data, Nonlinear pricing

JEL Classification: C13, D01, D11
AMS Classification: 91B16

1 Introduction

Elasticity of substitution is one of the most important cornerstones of consumer and firm theory in economics. It allows to study complementarity and substitutability among products and thus helps to establish the relationship between goods purchased by consumers. Note that, throughout this paper, substitutes and complements refer to the properties of the utility function (i.e., the curvature), not other derived concepts such as "p" and "q" substitutes [5]. In this context, the elasticity of substitution provides means to analyze the degree to which various inputs, such as consumption goods or labor inputs, can be substituted one for another [2] [8].

The curvature of the utility function, described by elasticity of substitution, is the primary, causative, and defining factor of substitutability or complementarity. All other measures of complementarity and substitutability thus might be compared with the definition of elasticity of substitution.

This paper investigates the concept of elasticity of substitution in consumer theory and explores various challenges associated with its estimation. The approach used in this study primarily relies on simulations, supplemented by illustrative examples of related theorems.

2 Elasticity of Substitution: Definition

The elasticity itself is defined as the ratio of percentage changes between two variables, meaning it measures the percentage change in one variable resulting from a percentage change in another variable. In this context, the elasticity of substitution refers to the percentage change in the ratio of goods purchased in response to a percentage change in the marginal rate of substitution [8]. Therefore, the elasticity of substitution can be written as:

$$\sigma_{ij} = \frac{\frac{\partial(x_j/x_i)}{x_j/x_i}}{\frac{\partial MRS_{ij}}{MRS_{ij}}} \tag{1}$$

Where $MRS_{ij} = MU_i/MU_j$ denotes marginal rate of substitution representing the slope of the indifference curve, which itself is a contour of the projection of utility function into two dimensions for given level of utility, while $MU_i = \partial U/\partial x_i$ corresponds to marginal utility, and x_i denotes quantity of good i to be purchased and then consumed.

[1] Prague University of Economics and Business, Department of Econometrics, W. Churchill sq. 1938/4, Prague, 130 67, The Czech Republic, petr.krautwurm@vse.cz
[2] Prague University of Economics and Business, Department of Econometrics, W. Churchill sq. 1938/4, Prague, 130 67, The Czech Republic, cernym@vse.cz

Note that, the elasticity of substitution can be written in logarithmic terms such as:

$$\sigma_{ij} = \frac{\partial \ln(x_j/x_i)}{\partial \ln(MU_i/MU_j)} \tag{2}$$

3 The General Consumer Optimization Problem

To illustrate the use of elasticity of substitution in consumer theory, it is essential to introduce the basic optimization problem of a consumer. The consumer purchases n goods at a shop, each with a price denoted by P_i, while being endowed with a fixed amount of money represented by M. Each good x_i is assumed to be a continuous variable. This optimization problem can be expressed in the following way:

$$\max_{\boldsymbol{x}} U(\boldsymbol{x}) \qquad s.t. \qquad \sum_{i}^{n} P_i x_i \leq M$$

Assume the utility function $U(\boldsymbol{x})$ to be well-behaved such that this problem is a case of convex optimization with binding constraint. Under this assumption, the problem might be solved via Lagrangian, which results in the following condition:

$$\frac{MU_i}{MU_j} = \frac{P_i}{P_j} \tag{3}$$

Where $MU_i = \partial U(\boldsymbol{x})/\partial x_i$. This condition is also being denoted as the Gossen's second law [10]. It is the most essential part in estimation of elasticity of substitution.

4 CES Utility Function

The last critical step necessary for the econometric analysis of the elasticity of substitution is to introduce a particular type of utility function with a constant elasticity of substitution (i.e., CES function). The CES function is defined as follows [1]:

$$U(\boldsymbol{x}) = \left(\sum_{i}^{n} a_i x_i^{\rho} \right)^{\frac{1}{\rho}} \tag{4}$$

Where $x_i \geq 0$ denotes the quantity of good i to be purchased and consumed, a_i denotes exogenous taste parameters, while the parameter $\rho \in (-\infty; 1)$ corresponds to the substitutability between products.

The importance of this utility function stands out when its elasticity of substitution is analyzed. Applying equation (1) for the elasticity of substitution to this particular utility function yields the following result:

$$\sigma_{ij} = \frac{1}{1-\rho}$$

The significance of the CES function thus lies in the fact that the elasticity of substitution for given pair of goods remains constant, regardless of the combination of quantities demanded [7] [6]. This implies that the CES function could be utilized to precisely describe the relationship between goods, thus identifying them as complements or substitutes. Furthermore, the CES utility function can identify the degree to which goods are complementary or substitutable [1].

Given that utility functions serve as non-unique descriptors of rational preferences, for any combination of $n-2$ fixed product quantities, the optimal consumption of the remaining products can be, in case of economic goods, approximated as if it was generated by a CES function, thereby identifying the precise relationship between goods in the case of this current optimum. However, it is crucial to emphasize that this approach relies on the assumption that each indifference curve is monotonically decreasing.

5 Elasticity of Substitution: Regression

In economics, the utility function is often considered a latent factor that cannot be directly observed. As a result, estimating the elasticity of substitution, as defined by equation (1), may appear to be an impossible task. However, it is essential to consider the complementary information about the consumer, concretely the consumer's optimal market behavior and the resulting relationship between marginal utilities and prices, which is outlined in equation (3). Since $MRS_{ij} = MU_i/MU_j$, the ratio of prices from equation (3) can be substituted into equation (1) resulting in:

$$\sigma_{ij} = \frac{\frac{\partial(x_j/x_i)}{x_j/x_i}}{\frac{\partial(P_i/P_j)}{P_i/P_j}} \tag{5}$$

Since x_i represents the goods that are purchased for consumption and P_i represents relevant prices, this equation now operates with observable variables. Now, it can be estimated by the following regression form:

$$ln\left(\frac{x_j}{x_i}\right) = \sigma_{ij} * ln\left(\frac{P_i}{P_j}\right) + H(\cdot) + \epsilon \tag{6}$$

Where $H(\cdot)$ is a function constant with respect to x_i, x_j, P_i, and P_j. This regression form, in turn, might be applied to a data set, containing transaction data of S consumers, each purchasing n products, in order to determine the average elasticity of substitution between a single pair of products. In case of estimating multiple elasticities of substitution between all pairs of products, this leads to $\binom{n}{2}$ regressions.

Because of this adjustment, the estimator can be applied to data resulting from market equilibrium without the need to estimate the demand first. Despite serving as a measure of consumer behavior, it is important to note that modifying the original elasticity of substitution formula to account for Gossen's law (3) links both sides of the market, namely supply and demand. This is important because market equilibrium data are typically the most accessible data, such as transaction data from receipts. This means that in order to use elasticity of substitution, it is sufficient for data to contain only the information on the quantity purchased and the prices consumers face. This fact stands out when comparing elasticity of substitution with other methods for classifying goods as complements or substitutes, such as cross-price elasticity of demand, which requires the demand to be estimated first [3]. What concerns the estimator itself, if used properly, it is consistent and does not differ in characteristics from ordinary least squares method.

5.1 Key Assumptions

As mentioned previously, the estimator's consistency depends on proper usage and certain assumptions. It measures the curvature of the indifference curve under the assumption that this curvature is constant. However, this is often not the case, and even in theory, no utility function can exhibit constant curvature in every direction [9]. As a result, the estimator can only offer a local approximation of the elasticity of substitution. For a given interval of quantities purchased, the appropriate use of the estimator is contingent upon the utility functions' curvature changing only slightly (i.e., not significantly).

However, another latent assumption is associated with how the MRS_{ij} was initially derived in the multivariate case ($n \geq 3$). The MRS_{ij} emerges from applying the total differential to the utility function, holding all irrelevant variables and utility level constant, for example, in case of $x_i \equiv x_1$ and $x_j \equiv x_2$:

$$dU|_{dU=0} = \frac{\partial U(x_1, x_2, \overline{x_3}, \ldots \overline{x_n})}{\partial x_1} dx_1 + \frac{\partial U(x_1, x_2, \overline{x_3}, \ldots \overline{x_n})}{\partial x_2} dx_2$$

Then, MRS_{ij} can be found by adjusting the resulting equation for $-dx_2/dx_1 = MU_1/MU_2$. This implies that the proper use of elasticity of substitution assumes the quantity of other irrelevant goods to remain fixed. When measuring the elasticity of substitution between all pairs of goods, this cannot be satisfied by definition.

Moreover, the estimator's applicability is related to the optimization problem being addressed. The estimator employs Gossen's law (3), derived for an optimization problem involving a binding constraint and linear prices. If consumers in the data solve a different type of problem, for example when encountering goods with quantity discounts or other types of nonlinear pricing, the estimator exhibits systematic bias.

Finally, the estimator's own limitations become apparent when the quantity of one good equals zero. This can arise from various reasons, such as the occurrence of perfect substitutes in the data or the good being an economic bad or neutral to the consumer. The estimator lacks the ability to differentiate between these possibilities.

Based on these considerations, four key assumptions can be established:

- **Assumption 1:** Utility functions for all consumers in the data exhibit approximately constant elasticity of substitution for given variations in quantities purchased.
- **Assumption 2:** The quantity of irrelevant goods remain constant.
- **Assumption 3:** Consumers face linear and binding constraint, meaning the data do not include goods subject to quantity discounts or any other form of nonlinear pricing.
- **Assumption 4:** The data contain only positive values.

The first assumption is the most critical one and cannot be relaxed. The rest can be relaxed to some extent, as shown in the following chapters.

6 Multivariate Elasticity of Substitution

In previous years, numerous generalizations of the concept of elasticity of substitution, suitable for functions with three or more variables, have been proposed [2]. However, these extensions are not always necessary, as the primary estimator (6) can be adjusted for changes in elasticity of substitution resulting from changes in irrelevant goods that were assumed to be constant, and thus provide optimal results.

Consider a simple utility function such as this:

$$U = \alpha x_1^2 + x_1 x_2 x_3$$

It can be shown that elasticity of substitution (1) between x_2 and x_3 is constant and equal to 1, while for the pair of x_1 and x_2 it is equal to the following:

$$\sigma_{12} = 2\frac{\alpha}{x_3}\frac{x_1}{x_2} + 1$$

In this scenario, the estimator (6) will exhibit systematic bias, unless accounted for the changes caused by variation in the variable x_3. However, if changes of elasticity of substitution due to changes in variables are included in regression under the function $H(\cdot)$ by controlling for levels of these irrelevant variables, the estimator starts to provide consistent results. This leads to the following modification of the estimator:

$$ln\left(\frac{x_j}{x_i}\right) = \sigma_{ij} * ln\left(\frac{P_i}{P_j}\right) + \sum_{k\neq i,j}^{n} \beta_k x_k + \epsilon \tag{7}$$

Where β_k denotes linear regression parameters and x_k represents irrelevant variables affecting the elasticity of substitution between x_i and x_j. Accounting for interferences from irrelevant variables, this adjustment allows to observe the clear effect of price ratio on product ratio, thus estimate the elasticity of substitution consistently. Note that, this functional form satisfies the condition for adjusted elasticity of substitution (5). Nonetheless, further examination of the estimator is warranted, particularly for more complex utility functions.

7 Nonlinear Budget Constraints

One of the primary issues with elasticity of substitution arises from nonlinear prices. However, this problem can potentially be resolved. The elasticity of substitution utilizes Gossen's law (3), derived from the general consumer optimization problem, which is then substituted for the marginal rate of substitution (5). If prices are nonlinear, such as in the case of quantity discounts, this becomes unfeasible and different substitution is required. To illustrate this, consider an optimization problem with prices determined by a general pricing function that depends on the quantity of the relevant good:

$$\max_{\boldsymbol{x}} U(\boldsymbol{x}) \qquad s.t. \qquad \sum_{i}^{n} P_i(x_i) \cdot x_i \leq M$$

Assume this pricing function to be well-behaved and thus at least twice differentiable and leading to binding constraint. In that case, solving this problem via Lagrangian method leads to the following solution:

$$\frac{MU_i}{MU_j} = \frac{P_i(x_i) + P_i'(x_i) \cdot x_i}{P_j(x_j) + P_j'(x_j) \cdot x_j} \tag{8}$$

This solution can be substituted for the marginal rate of substitution (5) in the same manner as Gossen's law (3). However, this approach introduces a new challenge: determining whether various types of quantity discounts in a store, such as lower prices for larger packages or direct discounts like two products for the price of one, can be approximated by the pricing function. Nevertheless, under the assumption that the pricing function is known, new prices per quantities purchased can be established, and the elasticity of substitution can be estimated. Yet, the question of properly approximating nonlinear pricing function remains a subject to subsequent research.

8 Perfect Substitutes and Economic Bads

A distinct issue arises from the presence of zero values in the data, both for prices and for goods. Zero prices are relatively rare, and they typically necessitate an experimental approach [4], rendering them problematic *per se*. Zero quantities, on the other hand, are more common in stores, often resulting from perfect substitutes, economic bads, or the inherent discreteness of shopping situations.

Fortunately, this issue has a straightforward solution. The estimator measures the average constant elasticity of substitution among consumers. From a theoretical standpoint, the presence of perfect substitutes leads to immeasurable elasticity of substitution anyway, as it becomes infinite in such cases. Concerning economic bads, no substitution is theoretically possible among them either. Thus, it can be concluded that the elasticity of substitution can be measured solely for consumers who purchase both goods, while also computing their ratio to the total number of consumers.

This approach, however, raises another concern when comparing single elasticities of substitution among different pairs of goods. It might imply that the average consumer has positive elasticities of substitution for different product pairs, while in reality, there may not be a single customer considering substitution between these items. Further research is necessary to identify the appropriate methodology for addressing this issue.

9 Conclusion

This paper explores several critical aspects of the elasticity of substitution estimator, highlighting its advantages, challenges, and outlining potential solutions for these aforementioned issues. The paper's contribution resides in establishing four assumptions concerning the consumer optimization problem that must be met for the standard elasticity of substitution estimator to yield consistent results, and in showing possible adjustments that have to be made in order to relax three of these assumptions.

Future research could concentrate on identifying more complex functions with analytically tractable elasticity of substitution and examining the estimator's properties within these contexts, ideally including non-linear constraints as well. Subsequently, it would be valuable to apply the theoretical findings of this paper to real-world data in subsequent investigations.

Acknowledgements

This work was supported by The Internal Grant Agency of Prague University of Economics and Business [VŠE IGS F4/52/2023]

References

[1] Arrow, K. J., Chenery, H. B., Minhas, B. S. B. & Solow, R. M. (1961). Capital-Labor Substitution and Economic Efficiency. *The Review of Economics and Statistics*, 43(3), 225-250.

[2] Blackorby, C. & Russell, R. R. (1989). Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities). *The American Economic Review*, 79(4), 882–888.

[3] Deaton, A. (1987). Estimation of own- and cross-price elasticities from household survey data. *Journal of Econometrics*, 36(1–2), 7–30.

[4] Hanley, N., Shaw, W. D. & Wright, R. E. (2003). *The New Economics of Outdoor Recreation*. Edward Elgar Publishing.

[5] Hicks, J. (1970). Elasticity of substitution again: substitutes and complements. *Oxford Economic Papers-New Series*, 22(3), 289–296.

[6] Ioan, C. A. & Ioan, G. (2011). A generalization of a class of production functions. *Applied Economics Letters*, 18(18), 1777–1784.

[7] McFadden, D. (1963). Constant Elasticity of Substitution Production Functions. *The Review of Economic Studies*, 30(2), 73-83.

[8] Stern, D. I. (2011). Elasticities of substitution and complementarity. *Journal of Productivity Analysis*, 36(1), 79–89.

[9] Uzawa, H. (1962). Production Functions with Constant Elasticities of Substitution. *The Review of Economic Studies*, 29(4), 291-299.

[10] Van Daal, J. (1996). From Utilitarianism to Hedonism: Gossen, Jevons And Walras. *Journal of the History of Economic Thought*, 18(2), 271–286.

Towards an Alternative Generalization of CES Function

Petr Krautwurm[1], Michal Černý[2]

Abstract. This paper explores ways to improve the usefulness of the Constant Elasticity of Substitution (CES) function in microeconomics. Although the concept of elasticity of substitution is a valuable tool for identifying substitutes and complements, the limitations of standard CES function, such as attributing the same elasticity of substitution to each pair of goods, have hindered its effectiveness. To address this issue, this article reviews previous research on the generalization of CES function. It examines the Nested CES function, representing the best possible generalization of CES function, and its implications for analyzing relationships between distinct product pairs. Subsequently, the paper proposes a new function, the Almost Constant Elasticity of Substitution (ACES) function, as another potential instrument for such analysis, and demonstrates its possible impact on the analysis of cross-price elasticity.

Keywords: CES function, Elasticity of substitution, Generalization, Nested CES function, Cross-price elasticity

JEL Classification: C13, D01, D11
AMS Classification: 91B16

1 Introduction

The Constant Elasticity of Substitution (CES) function is one of the most significant objective functions in economic theory [10]. Despite its widespread use and numerous important economic implications, it suffers from several critical shortcomings that significantly reduce its effectiveness. Most notably, the CES function assumes a constant elasticity of substitution for all product pairs. Since elasticity of substitution is a valuable tool in the analysis of substitutes and complements, this property limits its full potential. Therefore, this paper investigates possible ways to generalize the CES function, allowing for different yet constant elasticities of substitution between distinct product pairs. Furthermore, the paper proposes a search for a new function satisfying several assumptions that result in varying elasticities of substitution between different product pairs.

2 Generalizing CES Function

The CES function is an objective function, defined as follows:

$$U(\boldsymbol{x}) = \left(\sum_{i}^{n} a_i x_i^{\rho} \right)^{\frac{1}{\rho}} \tag{1}$$

Where $x_i \geq 0$ denotes quantity of good i that is purchased for consumption, a_i are taste parameters, and the parameter $\rho \in (-\infty; 1)$ corresponds to the substitutability between products. This can be shown by applying the common Robinson's [8] formula for elasticity of substitution [3] [11] on this particular function, which yields the following results:

$$\sigma_{ij} = \frac{\frac{\partial (x_j/x_i)}{x_j/x_i}}{\frac{\partial MRS_{ij}}{MRS_{ij}}} = \frac{MRS_{ij}}{x_j/x_i} \frac{1}{\frac{\partial MRS_{ij}}{\partial (x_j/x_i)}} = \frac{1}{1-\rho} \tag{2}$$

Where $\sigma_{ij} \in (0; \infty)$ is the elasticity of substitution, $MRS_{ij} = MU_i/MU_j$ denotes the marginal rate of substitution, and $MU_i = \partial U/\partial x_i$ corresponds to marginal utility.

The CES function's significance lies in its ability to accurately determine the relationship between products through the appropriate selection of parameters, while maintaining this relationship constant. This implies that two products

[1] Prague University of Economics and Business, Department of Econometrics, W. Churchill sq. 1938/4, Prague, 130 67, The Czech Republic, petr.krautwurm@vse.cz
[2] Prague University of Economics and Business, Department of Econometrics, W. Churchill sq. 1938/4, Prague, 130 67, The Czech Republic, cernym@vse.cz

can be designated as either substitutes or complements, and their relationship will remain unchanged, irrespective of any other factors. Nevertheless, except these advantages, the CES function has a notable limitation: the constant relationship applies uniformly to all pairs of products. This raises the question of whether a CES function generalization could exist that allows for different pairs of goods to exhibit distinct, yet constant, elasticities of substitution. This would entail maintaining a constant curvature of indifference curves, independent of unrelated goods. Regrettably, Uzawa [12] demonstrated that such a function cannot exist. The only feasible CES function generalization is the Nested CES approach [9] [2].

2.1 Composite Goods

Before analyzing the Nested CES function, it is essential to introduce the concept of composite goods. Composite goods represent product categories or bundles (e.g., food, clothing, or tools) rather than individual products themselves. These categories, though, are usually consumer-dependent and are not restricted to predefined classifications as given in the example. The only assumption relevant to composite goods is that they are either additively separable in the utility function or their relative prices are parallel [5] [1], which itself could be to some extent relaxed [6] [7]. To illustrate how composite goods work, consider the following consumer optimization problem:

$$\max_{x_1,y} U(x_1, y) = x_1 \cdot y \qquad s.t. \qquad P_1 x_1 + P_y y \leq M$$

Where $y = x_2 \cdot x_3 + x_3$ represents the composite good, and P_y is a suitable price index for it. The composite commodity theorem states that the consumer will achieve the same optimal outcome whether solving the problem with the composite good or the following with primary goods only:

$$\max_{x_1,x_2,x_3} U(\boldsymbol{x}) = x_1 \cdot (x_2 x_3 + x_3) \qquad s.t. \qquad P_1 x_1 + P_2 x_2 + P_3 x_3 \leq M$$

The primary advantage of this approach is that it facilitates a hierarchical analysis of the problem. Initially, consumers determine the optimal amount of composite good y to be consumed. Subsequently, they can identify the composition of the chosen good.

2.2 Nested CES Function

Incorporating composite goods is crucial when working with the generalization of the CES function, as it serves as the foundational component for constructing the Nested CES function. Consider the following utility function:

$$U(\boldsymbol{y}) = \left(\alpha_1 y_1^{\rho} + \alpha_2 y_2^{\rho}\right)^{\frac{1}{\rho}}$$

Where $y_1 = y_1(x_1, x_2)$ and $y_2 = y_2(x_3, x_4)$ are composite goods, composed of another CES function, resulting in a following scenario:

$$U(\boldsymbol{x}) = \left(a_1 \left(b_1 x_1^{\delta} + b_2 x_2^{\delta}\right)^{\frac{\rho}{\delta}} + a_2 \left(b_3 x_3^{\gamma} + b_4 x_4^{\gamma}\right)^{\frac{\rho}{\gamma}}\right)^{\frac{1}{\rho}} \tag{3}$$

In this scenario, three constant elasticities of substitution can be distinguished: the first for the pair of x_1 and x_2, the second for x_3 and x_4, and the third for y_1 and y_2. This aligns with Uzawa's (1962) findings, which demonstrated that the only possible generalization of the CES function is such that the utility function consists of at most $S \leq n/2$ product groups. Within each group, products exhibit constant and similar elasticity of substitution, while products between groups display varying ones. Consequently, it is possible to categorize the marginal rate of substitution $MRS_{ij} = MU_i/MU_j$ and the elasticity of substitution into intra and extra group classifications. This is essential, as the process of deriving the elasticity of substitution (2) depends on being able to identify the marginal rate of substitution, and understanding its functional form can assist in determining whether the included differentiation is analytically tractable. The results can be illustrated by comparing the outcomes when analyzing, for example, the combination of x_1 with x_2, representing the intra group, versus the combination of x_1 with x_3, representing the extra group:

$$MRS_{intra} = \frac{\partial U(\boldsymbol{x})/\partial x_1}{\partial U(\boldsymbol{x})/\partial x_2} = \frac{b_1}{b_2}\left(\frac{x_1}{x_2}\right)^{\delta-1} \qquad \Longrightarrow \qquad \sigma_{intra} = \frac{1}{1-\delta} \tag{4}$$

$$MRS_{extra} = \frac{\partial U(\boldsymbol{x})/\partial x_1}{\partial U(\boldsymbol{x})/\partial x_3} = \frac{a_1\left(b_1 x_1^{\delta} + b_2 x_2^{\delta}\right)^{\frac{\rho}{\delta}-1}}{a_2\left(b_3 x_3^{\gamma} + b_4 x_4^{\gamma}\right)^{\frac{\rho}{\gamma}-1}}\left(\frac{b_1 x_1^{\delta-1}}{b_3 x_3^{\gamma-1}}\right) = \left(\frac{a_1 y_1^{\rho-\delta}}{a_2 y_2^{\rho-\gamma}}\right)\left(\frac{b_1 x_1^{\delta-1}}{b_3 x_3^{\gamma-1}}\right) \tag{5}$$

As can be observed, the elasticity of substitution applied to the former combination (4) yields the same result as the CES function (2), while this does not hold true for the latter (5). The elasticity of substitution between x_1 and x_3 depends not only on the goods' ratio but also on the changes in these goods caused by alterations in the ratio itself.

Despite the noticeable limitations of this generalization, it still offers some benefits. One advantage is the utility function's capacity to establish a constant relationship between composite goods themselves. This is valuable because it enables to determine the substitutability between products from two different groups via the relationship of their respective composite goods. By suitably comparing this behavior with the CES function, it is possible to assess, to a certain extent, whether it corresponds more closely to either complements or substitutes.

An additional advantage is the capacity to measure the elasticity of substitution for categories directly from data. If, in data with n goods, several product groups emerge with statistically indistinguishable elasticities within groups but differing across groups, composite goods can be constructed based on the Nested CES function. Following this, the elasticity of substitution between these groups can be then measured and compared with the corresponding extra-group elasticities.

2.3 ACES Function Proposal

The inability to construct a fully generalized CES function suggests that all alternative generalizing functions must exhibit some variability in the elasticity of substitution to a certain extent. A critical question that arises is determining the minimal possible variability in the elasticity of substitution such that, for reasonable changes in goods, the elasticity of substitution between pairs effectively remains constant. To illustrate this, assume that the elasticity of substitution between any pair of goods takes the following form:

$$\sigma_{i,j} = \sigma^*_{i,j} \cdot g(\boldsymbol{x}) + h(\boldsymbol{x})$$

Where $\sigma^*_{i,j}$ is a scalar constant, dependent solely on parameters of the utility function, while $g(\boldsymbol{x})$ and $h(\boldsymbol{x})$ are functions of variables possessing the following property:

$$g(\boldsymbol{x}) \cong 1 \quad \& \quad h(\boldsymbol{x}) \cong 0 \qquad [\boldsymbol{x} \in \mathbb{A}]$$

Where $\mathbb{A}$ represents a reasonably broad set of product combinations. This reasonability may be characterized by a multidimensional cone from the initial position, ensuring that the function exhibits approximately constant elasticity of substitution for the most common variations in variables.

The importance of this assumption resides in the fact that within the set $\mathbb{A}$, every pair of goods can be locally approximated by their own CES function. This function provides a constant elasticity of substitution for any given quantity of other irrelevant goods within the set $\mathbb{A}$. This holds particular relevance as Robinson's elasticity of substitution (2) serves to measure the curvature of the indifference curve, thus capturing the interrelationship between goods.

Although it remains impossible to create a generalization of the CES function that allows for a different yet constant elasticity of substitution between varying pairs of goods, the proposed approach enables local approximation of the original utility function as if this condition was satisfied.

Moreover, this function should also satisfy an additional property such that its demand functions are analytically tractable. However, this requirement may potentially conflict with the previous condition, implying that multiple versions of the Almost Constant Elasticity of Substitution (ACES) function could be viable. Thorough examination of the appropriate functional form is necessary in future research.

3 Implications of Generalizing CES Function

The inability to construct a fully generalized CES function reveals significant insights about the relationships within a complete set of products. Uzawa's [12] findings indicate that the elasticity of substitution between two goods cannot be, in general, independent of the elasticity of substitution between another pair of goods, where one good is the same as in the previous pair and the other is different. Consequently, the generalization of the CES function could offer a means to analyze this measure of co-substitutability, thereby examining the dependency of the relationship between two goods on their relationships with all other goods.

Another crucial implication lies in the fact that other measures of substitutability or complementarity depend on the elasticity of substitution. As such, it could be valuable to investigate how another famous measure, the cross-price elasticity, is influenced by the relationship between examined goods and intermediary ones.

3.1 Nested CES and Special Cases

Nested CES (3), as defined previously, consists of a tree of composite goods with each of these determined by its own CES function. As a result, the restrictions on parameters $\rho_k \in \langle -\infty, 1\rangle$ are the same at each level of the Nested CES. In this context, Lagomarsino [4] specifies various suitable nesting structures for three- and four-input Nested CES functions. However, it could be argued that another common structure in economic theory was overlooked in this analysis: the case where the Nested CES is constructed by composite goods that both depend on the same primary good.

Consider a modification to the previous case (3), in which both composite goods share the same primary good: $y_1 = y_1(x_1, x_2)$ and $y_2 = y_2(x_2, x_3)$. This would lead to the following Nested CES function:

$$U = \left(a_1 \left(b_1 x_1^{\delta} + b_2 x_2^{\delta}\right)^{\frac{\rho}{\delta}} + a_2 \left(b_3 x_2^{\gamma} + b_4 x_3^{\gamma}\right)^{\frac{\rho}{\gamma}}\right)^{\frac{1}{\rho}} \tag{6}$$

Now, consider the special cases of standard CES function, which might be applied also to the Nested CES function.

$$\left(a_1 y_1^{\rho} + a_2 y_2^{\rho}\right)^{\frac{1}{\rho}} \approx \begin{cases} MIN\{a_1 y_1, a_2 y_2\} & [\rho = -\infty] \\ y_1^{a_1} y_2^{a_2} & [\rho = 0] \\ a_1 y_1 + a_2 y_2 & [\rho = 1] \end{cases}$$

The Nested CES with a common primary good (6) allows for 18 effectively unique combinations of special cases. A thorough examination of these cases reveals that certain utility function shapes are unattainable, meaning that the elasticity of substitution for specific product pairs constrains the values another elasticity of substitution can take. For instance, no utility function exists with $\sigma_{1,2} = 0$, $\sigma_{2,3} = \infty$, and $\sigma_{1,3} = 1$.

This finding could have significant implications for investigating consumer rationality. If rational preferences (i.e., those describable by utility function) determine which combinations of elasticities of substitution cannot exist, it may be possible to measure whether these unattainable patterns emerge in consumer behavior. Further research is required to explore this topic.

3.2 Cross-Price Elasticity

Elasticity of substitution is not the only means of identifying goods as substitutes or complements. An alternative approach consists of estimating cross-price elasticity of demand. This elasticity measures the percentage change in marshallian demand for certain good x_i^* in response to percentage change in price of different good:

$$XPE = \frac{\partial x_i^*}{\partial P_j} \frac{P_j}{x_i^*} \tag{7}$$

Applying this formula to the demand derived from CES function yields the following result:

$$XPE = \frac{\partial x_i^*}{\partial P_j} \frac{P_j}{x_i^*} = -\frac{\rho}{\rho - 1} \cdot K \tag{8}$$

Where $K = K(P_1, \ldots, P_n, a_1, \ldots, a_n)$ is a function depending solely on prices and taste parameters. Since $1/(\rho - 1)$ will be always negative for $\rho \in (-\infty, 1\rangle$, which ensures convex indifference curve, the sign of the whole cross-price elasticity depend only on the sign of ρ.

- If $\rho > 0$, goods are substitutes, and $XPE > 0$.
- If $\rho < 0$, goods are complements, and $XPE < 0$.

Based on these results, it would be fruitful to investigate the behavior of cross-price elasticity of demand when the utility function allows for different elasticities of substitution between various pairs of goods. This is a potential application for the ACES function.

4 Conclusion

In conclusion, this paper has explored the possibilities of generalizing the CES function to allow for distinct yet constant elasticities of substitution between different product pairs. Although the findings reveal limitations in fully generalizing the CES function, the Nested CES function, and the proposed ACES function emerge as potential avenues for further investigation. These approaches could provide valuable insights into the relationships between goods and shed light on the implications of consumer rationality.

Moreover, the paper has discussed the relationship between cross-price elasticity of demand and elasticity of substitution, emphasizing the importance of understanding the interplay between these measures. As a result, future research should focus on identifying suitable generalizations of CES functions that maintain analytical tractability and are capable of capturing the unique characteristics of varying elasticities of substitution between distinct product pairs.

Acknowledgements

This work was supported by The Internal Grant Agency of Prague University of Economics and Business [VŠE IGS F4/52/2023]

References

[1] Carter, M. (1995). An expository note on the composite commodity theorem. *Economic Theory*, 5(1), 175–179.

[2] Keller, W. J. (1976). A nested CES-type utility function and its demand and price-index functions. *European Economic Review*, 7(2), 175–186.

[3] Knoblach, M. & Stöckl, F. (2020).What determines the elasticity of substitution between capital and labor? A literature review. *Journal of Economic Surveys*, 34(4), 847–875.

[4] Lagomarsino. (2021). Which nesting structure for the CES? A new selection approach based on input separability. *Economic Modelling*, 102, 105562.

[5] Leontief, W. (1936). Composite Commodities and the Problem of Index Numbers. *Econometrica*, 4(1), 39.

[6] Lewbel, A. (1996). Aggregation without Separability: A Generalized Composite Commodity Theorem. *The American Economic Review*, 86(3), 524–543.

[7] Reed, A. J., Levedahl, J. W. & Hallahan, C. (2005). The Generalized Composite Commodity Theorem and Food Demand Estimation. *American Journal of Agricultural Economics*, 87(1), 28–37.

[8] Robinson, J. (1933) *The Economics of Imperfect Competition (1st edn)*. London: Macmillan.

[9] Sato, K. (1967). A Two-Level Constant-Elasticity-of-Substitution Production Function. *The Review of Economic Studies*, 34(2), 201-218.

[10] Sato, R. (1975). The Most General Class of CES Functions. *Econometrica*, 43(5/6), 999-1003.

[11] Stern, D. I. (2011). Elasticities of substitution and complementarity. *Journal of Productivity Analysis*, 36(1), 79–89.

[12] Uzawa, H. (1962). Production Functions with Constant Elasticities of Substitution. *The Review of Economic Studies*, 29(4), 291-299.

Energy Consumption and Economic Growth in the Czech Republic and Slovakia

Radmila Krkošková[1]

Abstract. Energy consumption and economic growth are interconnected, and the relationship between them is complex and dependent on various factors. Key aspects of this relationship include: energy consumption as a driver of economic growth; technological progress and efficiency; structural changes in the economy; policies and regulations.

This article examines the long-term relationship between energy consumption and real GDP for the Czech Republic and Slovakia from 2005 to 2022. Many studies have explored the linkages between energy consumption, economic growth, and energy efficiency. The aim of this contribution is to contribute to this topic by analyzing the Granger causality between the indicators. This paper focuses on the predictive power (Granger causality) rather than estimating the true causal relationship of the VAR/VECM model. The results for both countries indicate that energy consumption Granger-causes GDP. This means that energy-saving policies may slow down the pace of GDP growth. These findings are based on the analysis of Eurostat data from 2005 to 2022 for the Czech Republic and Slovakia, using the statistical software EViews 11 for calculations.

Keywords: ADF test, energy, GDP, Granger causality, VEC model

JEL Classification: C22, Q43, Q48
AMS Classification: 62P20, 91B62

1 Introduction

This short article expands on the original work [14] and focuses on new data, presenting additional approaches to the researched issue.

Energy consumption and economic growth are closely related, as economic growth typically leads to increased energy consumption. As countries develop and their economies grow, they tend to consume more energy to fuel their industries, transportation systems, and homes. This relationship between energy consumption and economic growth is often referred to as the "energy-growth nexus." However, the extent to which energy consumption drives economic growth and vice versa is a matter of debate among economists and policy-makers.

On one hand, some argue that increased energy consumption is necessary for economic growth, as it provides the energy needed to power economic activity. This is particularly true in developing countries where energy infrastructure is often lacking and a lack of access to energy can limit economic development.

On the other hand, others argue that increased energy consumption can actually be a drag on economic growth, as it can lead to higher energy costs and decreased competitiveness. Additionally, the negative environmental impact of increased energy consumption can lead to economic costs in the form of environmental damage and the costs of mitigating and adapting to climate change.

It is natural that the dependence of the GDP and the energy consumption is not the same in the Czech Republic and Slovakia. The article contains the following sections: a review of the literature, the econometric methods used, the models for both countries, and conclusion.

[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

2 Literature Review and Data

2.1 Literature Review

There have been numerous articles written by scientists and researchers about the relationship between energy consumption and economic growth. One notable article is article [10] "Energy consumption and growth: a review of international empirical literature".

This paper focuses on a review of existing literature on the causal relationship between energy consumption and economic growth. There are currently four views regarding the direction of causality between energy consumption and economic growth in the literature. The first view posts on the notion that energy consumption Granger-causes economic growth. This view is noted as the "energy-led growth hypothesis". The second view, known as the "growth-led energy consumption hypothesis", supports the view that economic growth Granger-causes energy consumption. The third view argues for a bidirectional causal relationship between energy consumption and economic growth, whilst the fourth view argues for no causal relationship between the two variables.

Although other literature surveys have focused mainly on studies that use total energy consumption as a proxy for energy consumption, few have consolidated the literature based on different energy sources and how these impacts on the direction of causality. The findings from the literature reviewed in this study reveal that although there is no consensus yet on the direction of causality between energy consumption and economic growth, the conventional wisdom is in favour of the feedback hypothesis, where energy consumption and economic growth Granger-cause one another. In terms of energy sources, the study reveals that electricity energy consumption predominates the growth-led energy hypothesis, whilst renewable energy consumption predominates the feedback hypothesis.

The study [16] investigates the relationship between energy consumption, economic growth and financial development in India by using the annual data for the period 1971-2009. An application of Auto Regressive Distributed Lag (ARDL) approach to cointegration results suggest that energy consumption is positively and significantly impacted by proportion of urban population in total population, while the same is negatively and significantly impacted by financial development, economic growth and proportion of industrial output in total output.

The dependence on composition of industry and the relationship between GDP and energy consumption has been described in the article [7]. Authors state that there is not a clear causality. The article [19] presents literature review on this topic. There are analyses a larger number of countries and authors state that the causality between GDP and electricity consumption varies across countries. Authors divided their study [18] into two groups, the first group contained the low-income countries, and the second one contained the middle-income ones. The finding of the article is that the causality of low-income countries is such that GDP affects the amount of energy consumption, and it is a mutual causality in middle-income countries. The paper [15] shows that In China, the demand for energy is increasing as a result of improving people's living standards. Authors proved that there is a long-term stable relationship between energy consumption and economic growth, and there is unidirectional causality from economic growth to energy consumption. The situation about energy consumption in Brazil in the period from 1971 to 2010 is described in [2]. Authors state that there is a positive relationship between energy consumption and economic progress in Brazil.

Regarding panel data analysis the situation about energy consumption in Baltic countries in the period from 1992 to 2011 is described in [8]. Author state that there $GDP \rightarrow EC$. The same result we can find out in the article [20], which describes 5 South Asian countries during the period 1990-2014. The relationship $EC \leftrightarrow GDP$ was confirmed in the articles [1], [17].

The results of this article: causality energy consumption causes GDP was confirmed for both countries the Czech Republic and Slovakia.

2.2 Data and Methods

The data used have the character of quarterly time series in the period from 2005Q1 to 2022Q4. All values were considered in logarithmic terms. Data were obtained from the Eurostat database [5], [6]. Variables used in our research are: *GDP* and the energy consumption. Variables *CZ_GDP, SK_GDP* are listed in million units of national currency and considered in logarithmic terms. Variables *CZ_EC, SK_EC* are listed in million tonnes of oil equivalent (TOE) and considered in logarithmic terms.

The variables (*CR_GDP, CR_EC*) are type of I (1) as the Table 1 shows. Therefore, the long-run co-integration relationships may exist between these time series. Using the Johansen's method, it was confirmed the existence

of 1 co-integration relationship for VECM (1) in the case of the Czech Republic. The variables (*SK_GDP, SK_EC*) are zero-order integrated. In the case of Slovakia the energy consumption use Granger causes the GDP. If the variables are not co-integrated we can use the VAR model to determine the directions of causality.

The first step is testing the stationarity of variables included in the model or their first differences. Table 2 shows the test results for all variables of all countries. 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	Prob.	Result	Variable	n/c/c+t	T-stat	Prob.	Result
CR_GDP	c	−0.578	0.781	N	D(*CR_GDP*)	c	−4.951	0.001***	S
CR_EC	c	−1.465	0.621	N	D(*CR_EC*)	c	−4.654	0.001***	S
SK_GDP	c+t	−5.815	0.002***	S					
SK_EC	n	−4.233	0.037**	S					

Statistical significance at the 0.01 level (***), at the 0.05 level (**), at the 0.1 level (*)

Table 1 ADF test

The Dickey-Fuller test (ADF) was used to test the stationarity. Authors Engle and Granger [4], and Enders [3] argue that most time series in macroeconomics are non-stationary or integrated with order I (1).

In the article is performed the individual analysis of each country. We examine the long-run relationship in a function *GDP* = *f* (*EC*); where *GDP* = ln of real Gross Domestic Product; *EC* = ln of Energy Consumption. The methods are described in the articles Hendry & Juselius [12], [13].

If all the variables are stationary, then vector autoregressive (VAR) model can be used. If the variables of type I (1) we use the error correction model (VECM) and test the cointegration and determine the Granger causality connections.
The general form of the VECM model is:

$$\Delta y_t = \gamma \Delta x_t + \alpha(y_{t-1} + \beta x_{t-1}) + u_t \tag{1}$$

where x_t, y_t are economic variables, u_t is the residual of the variable, $\Delta x_t = x_t - x_{t-1}$, expressions y_{t-1} , x_{t-1} are error correction terms, parameter β describes long-term cointegration relationships between variables, parameter γ describes short-term relationships, and parameter α indicates the speed of adaptation to the equilibrium state.

In our case, two variables will be used and the lag lengths is equal to one. The VAR model can be written as follows:

$$\begin{aligned} x_t &= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 y_{t-1} \\ y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} \end{aligned} \tag{2}$$

3 Data Analysis

3.1 The Czech Republic

Existence of one long-term bond can be specified by a co-integration equation:

$$EQ\_CR = CR\_GDP_t + 26.46 CR\_EC_t \tag{3}$$

The co-integration vector, (1.00; 26.46), indicates that a 1% decrease in CR_EC (energy consumption) will result in a 26.46% increase in CR_GDP (Gross Domestic Product). Therefore, reducing energy consumption leads to faster GDP growth. This finding provides motivation for implementing "austerity programs."

Table 2 demonstrates the statistical significance of the correction term and confirms that the model explains the return to long-term equilibrium. Regarding the regression coefficients, it can be argued that there is a negative relationship between GDP and energy consumption, with quarterly delays.

Error Correction:	**D(*CR_GDP*)**	**D(*CR_EC*)**
CointEq1	-0.012**	-0.014***
D(*CR_GDP*(-1))	-0.077	0.191
D(*CR_EC*(-1))	-0.731***	-0.422***
C	0.014***	-0.0002
R-squared	0.62	0.28

Statistical significance at the 0.01 level (***),at the 0.05 level (**),at the 0.1 level (*)

Table 2 Estimates VECM(1)

This part of article deals with the testing of short-term relationships (Granger causality). Granger causality only provides information regarding forecasting ability and does not offer insight into the true causal relationship between variables, it can be found in [9], [11]. It is necessary to work with the stationary time series. Due to the fact that these are quarterly data, Granger causality is tested at the 1, 2, 3, 4 delay. The similar procedure is given in [21]. We consider the 5% significance level. The results of the series 1 delay test are shown in Table 3.

Null Hypothesis:	**Statistic**	**Sign.**
D(*CR_GDP*) does not Granger cause D(*CR_EC*)	2.31	0.152
D(*CR_EC*) does not Granger cause D(*CR_GDP*)	4.24	**0.069**

Table 3 Pairwise Granger causality tests (lag 1)

The Table 4 shows that the energy consumption use Granger causes the *GDP*. It means that the energy conservation policies can retard the growth rate of *GDP*.

3.2 Slovakia

The variables (*SK_GDP*, *SK_EC*) are zero-order integrated, i.e. I(0). It shows Table 1.

The VAR models can be written as

$$\begin{aligned} SK\_EC &= 0.894 + 0.261.SK\_EC(-1) - 0.031.SK\_GDP(-1) \\ SK\_GDP &= 1.121 - 0.542.SK\_EC(-1) + 0.984.SK\_GDP(-1) \end{aligned} \tag{4}$$

The first equation shows that the dependent variable *EC* is positively related to the rise in the *EC* with quarterly delays, and negatively related to the rise in the *GDP*, with quarterly delays. The Table 4 shows that the energy consumption use Granger causes the *GDP*. It means that the energy conservation policies can retard the growth rate of *GDP*.

Null Hypothesis:	**Statistic**	**Sign.**
SK_GDP does not Granger cause *SK_EC*	0.02	0.851
SK_EC does not Granger cause *SK_GDP*	8.12	**0.008**

Table 4 Pairwise Granger causality tests (lag 1)

The *SK_GDP* series is trend-stationary, so it could be count regression analysis (Table 5). According to regression coefficient we can state, that regression coefficient is significant at the level 10%.

Variable	**Coefficient**	**Std.error**	**t-ratio**	**Sign.**
SK_EC	-13254.6	7296.2	-1.82	0.08
C	208579.8	80370.1	2.59	0.01

Table 5 Regression analysis for Slovakia

3.3 Model Assumptions, Impulse Response Function

The model assumptions for all countries were satisfied, as the null hypotheses regarding autocorrelation, heteroscedasticity, and normality were not rejected at the 5% significance level. The residual component of the model meets the necessary requirements. In the case of the Czech Republic and Slovakia, it was found that energy use Granger causes GDP. Figure 1, presented in the first column, illustrates the response function to a positive exogenous shock in energy consumption. The pattern of the function is similar for both countries. Specifically, a positive exogenous unit shock in energy consumption initially leads to a decline in GDP in Q2, followed by an increase in Q3.

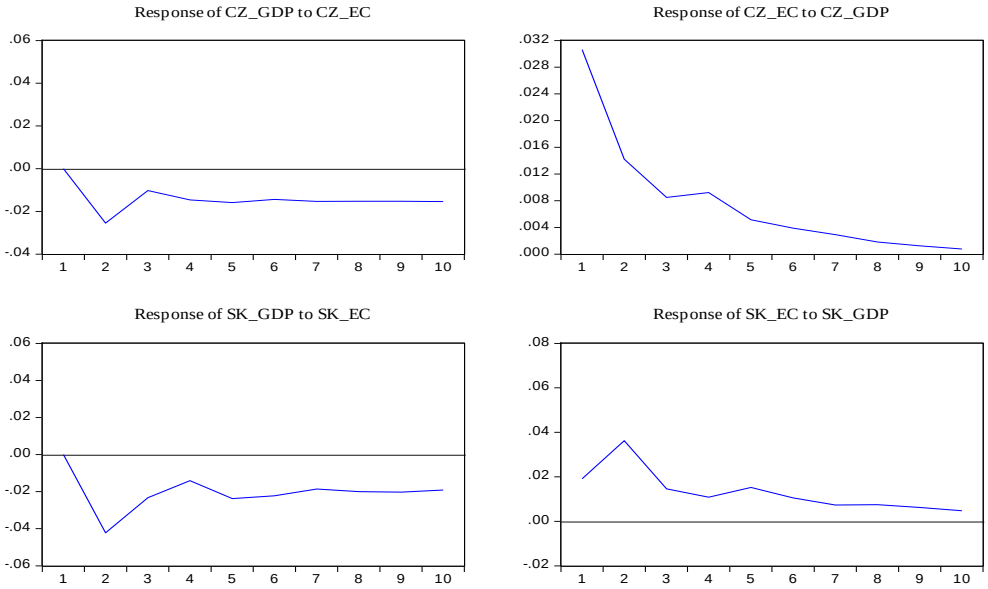


Figure 1 Response to Cholesky One S.D.Innovations

4 Conclusion

The Czech Republic and Slovakia, like many other countries, have seen a positive correlation between energy consumption and economic growth in recent decades until 2021. Since the 1990s, countries' energy consumption has steadily increased, contributing to the growth of their economies. In 2021, the situation has changed and energy consumption is decreasing due to the increase in energy prices on the market. According to the International Energy Agency, the Czech Republic's and Slovak's total energy consumption has increased by over 40% between 1990 and 2018. This increase was largely driven by the growth of the country's industrial sector, which has been a major contributor to the country's economic growth. However, in recent years, the both countries have been making efforts to increase its energy efficient and reduce their dependence on fossil fuels. The countries have set ambitious targets to reduce their greenhouse gas emissions and increase the share of renewable energy in their energy mix. For example, the Czech Republic has set a target to increase the share of renewable energy in its electricity generation mix to 25% by 2030. Slovakia has set a target to reduce its greenhouse gas emissions by 40% by 2030 compared to 1990 levels.

For the Czech Republic, the co-integration relationship was proved and in the next step, it was used the VECM model. A cointegration test was performed and this test proved the existence of a long-term equilibrium relationship between the variables. The results of the VECM model show that in the Czech Republic there is a short-term causal link from energy consumption to economic growth. These findings suggest that it is important not to reduce energy consumption, as this could lead to a reduction in economic prosperity.

In Slovakia, and the Czech Republic, it has been shown that energy consumption could lead to economic growth. The energy conservation issues should be addressed with the utmost caution, as these policies may have a negative impact on economic growth. In the case of Slovakia, there is a negative relationship and it means that decreasing the energy consumption leads to the faster increasing GDP. It is a motivation for economic measures. On the other hand, in the case of the Czech Republic there is positive relationships between energy consumption and economic development and it means that the economic measures slow down economic growth.

Energy consumption and economic growth will continue to be investigated through scientific research. For example, whether improving energy efficiency can support economic growth while reducing energy consumption. There is also a growing interest in renewable energy sources. Do they have an impact on economic growth? Consideration should also be given to tax incentives for energy efficiency, greenhouse gas emissions reduction, and support for innovation in the energy sector. These new research studies can contribute to a better understanding of the relationship between energy consumption and economic growth and can help in formulating policies and strategies for sustainable development.

All things considered, the relationship between energy consumption and economic growth is complex and depends on a range of factors, including the state of a country's economy, its level of development, and its energy policies.

Acknowledgements

This research was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2023.

References

[1] Alvarado, R., Ponce, P., Alvarado, R., Ponce, K., Huachizaca, V., & Toledo, E. (2019). Sustainable and non-sustainable energy and output in Latin America: A cointegration and causality approach with panel data. *Energy Strategy Reviews*, 26, 100-369.

[2] Aslam, N., Shahid, A.U., Rathore, M. & Tariq, M.I. (2014). An Empirical Analysis of Energy Consumption and Economic Growth in Brazil. *Journal of Economics and Development Studies*, 2(2), 591-599. [Online]. Available at: http://jedsnet.com/vol-2-no-2-june-2014-abstract-34-jeds [cited 2023-05-22].

[3] Enders, W. (2014). *Applied econometric time series*. Hoboken: Wiley.

[4] Engle, R. F. & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55 (2), 251-276. [Online]. Available at: https://www.jstor.org/stable/1913236 [cited 2023-04-21].

[5] Europa.eu, (2023a). Eurostat database. Available at: http://ec.europa.eu/eurostat/data/database [cited 2023-04-28].

[6] Europa.eu, (2023b). Eurostat database. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_statistics_-_an_overview#Energy_dependency [cited 2023-04-28].

[7] Faisal, T. & Resatoglu, N.G. (2016). Energy Consumption, Electricity, and GDP Causality; The Case of Russia, 1990-2011. *Procedia Economics and Finance*, 39, 653-659.

[8] Furuoka, F. (2017). Renewable electricity consumption and economic development: New findings from the Baltic countries. *Renewable and Sustainable Energy Reviews*, 71, 450-463.

[9] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica.* 37(3), 424–438.

[10] Gwenhure, Y. & Odhiambo, N. (2015). Energy consumption and growth: a review of international empirical literature. *Economics and Policy of Energy and the Environment*, 3, 47-70

[11] Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

[12] Hendry, D. & Juselius, K. (2000). Explaining Cointegration Analysis: Part I. *The Energy Journal*, 21(1), 1-42.

[13] Hendry, D. & Juselius, K. (2001). Explaining Cointegration Analysis: Part II. *The Energy Journal*,22(1),75-120.

[14] Krkošková, R. (2021). Causality between energy consumption and economic growth in the V4 countries. *Technological and Economic Development of Economy*, 27(4), 900-920. [Online]. Available at: https://journals.vilniustech.lt/index.php/TEDE/article/view/14863 [cited 2023-03-28].

[15] Li, X., Zhou, D. & Zhang, H. (2019). Quantitative analysis of energy consumption and economic growth in China. *IOP Conference Series Earth and Environmental Science*, 237 (4):042016.

[16] Mahalik, M. K. & Mallick, H. (2014). Energy consumption, economic growth and financial development: exporing the empirical linkages for India. *The Journal of Developing Areas*, 48(4), 139-159.

[17] Maji, I.K., Sulaiman, C., & Abdul-Rahim, A.S. (2019). Renewable energy consumption and economic growth nexus: A fresh evidence from West Africa. *Energy Reports*, 5, 384-392.

[18] Ozturk, I., Adslan, A. & Kalyoncu, H. (2010). Energy consumption and economic growth relationship: Evidence from panel data for low and middle income countries. *Energy Policy*, 38(8), 4422-4428.

[19] Payne, J.E. (2010). A survey of the electricity consumption-growth literature. *Applied Energy*, 87(3), 723-731.

[20] Rahman, M.M., & Velayutham, E. (2020). Renewable and non-renewable energy consumption-economic growth nexus: New evidence from South Asia. *Renewable Energy*, 147, 399-408.

[21] Stoklasová, R. (2018). Default rate in the Czech Republic depending on selected macroeconomic indicators. *E&M Economics and Management*, 21(2), 69-82. [Online]. Available at: https://dspace.tul.cz/items/4bc4189e-a2cc-4057-b27a-3d244c2a40c1 [cited 2023-04-21].

Efficiency Analysis of Building Material Producers in the Czech Republic

Martina Kuncová[1], Simona Činčalová[2], Petr Musil[3]

Abstract. The construction industry is one of the key sectors of the Czech economy and is often considered as an important indicator of economic development. It is a sector that is very sensitive to the economic cycle, but usually reacts with a delay to significant fluctuations in the cycle. Demand for construction work and building materials is influenced by the situation on the labour market and the development of household disposable incomes or interest rates for housing finance, as well as by the price of building land. The supply of construction materials is mainly influenced by the production capacity of firms, the prices of construction materials and works, or expected market regulations. The situation in the building materials market has not received much attention in the available literature. Therefore, in this paper, we focus on the largest producers of building materials in the Czech Republic and the development of the efficiency of these firms in the year 2018 based on 2 data sources and using DEA models. The aim is to analyse the efficiency of companies in the given year and to examine whether large and established companies performed better than small and medium-sized companies. At the same time, the aim is to see if we obtain similar results using different economic data from 2 different sources. DEA models with 3 inputs 1 output, resp. 1 input and 3 outputs, used on 2 sets of data showed that large companies are more efficient than small and medium companies in this sector.

Keywords: construction industry, building materials production, DEA models

JEL Classification: C44, L11, D24
AMS Classification: 90C08

1 Introduction

Construction is one of the key sectors of the Czech national economy. It is not only a significant consumer of industrial products, but also a field that affects the appearance of municipalities and cities, respectively the landscape. The construction industry is considered one of the important indicators of economic development. Together with engineering, it plays a decisive role in the investment construction.

The number of companies operating in the construction market in the Czech Republic is huge, as there are many areas they focus on, from the production of building materials to construction and reconstruction. In general, the construction industry includes companies focused on the construction of houses and apartments. In this article, however, we focus on firms producing building materials. According to the NACE classification of economic activities, these activities belong to area C - Manufacturing, and in particular to sections 2332 – Manufacture of bricks, tiles and construction products, in baked clay, and 2361 – Manufacture of concrete products for construction purpose [9]. In the construction industry, the 2008 crisis manifested itself over several years, e.g. the number of building permits issued for new buildings in the Czech Republic was declining until 2014, only after which there was some recovery and an extreme increase in 2018 [15]. Manufacturers immediately started to react to this situation. The two largest domestic brick manufacturers, Wienerberger and Heluz, increased production capacity by up to 15% year-on-year, while other manufacturers produced at the edge of their capacities. Even so, there was a temporary shortage of building materials on the Czech market. This situation persisted into 2019 (see Figure 1), and it was only the restrictions caused by the COVID pandemic that halted the development of the construction boom.

As there are not many publications devoted to efficiency analysis in the construction industry, we decided to use the cooperation with CEEC Research [13] to analyse the largest Czech companies in the field of building materials production and to subject the obtained data to DEA model analysis. Due to the large fluctuations in the sector during and after the pandemic years and non-availability of all necessary data, only the year 2018 is analysed in

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, kuncova@vspj.cz.
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, simona.cincalova@vspj.cz.
[3] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, petr.musil@vspj.cz.

this paper. The aim is to assess the efficiency of companies in 2018, including whether large and established companies performed better than small and medium-sized ones.

2 Construction Industry and the Czech Economy

Construction industry has a non-negligible share in the creation of gross value added (GVA) and gross domestic product (GDP), but also in employment. In the last ten years, it contributed between five and seven percent to the creation of GVA and made up between seven and ten percent of the GDP [8]. In recent years, however, the construction industry's share of both mentioned quantities has been slightly decreasing.

In terms of employment, the construction industry can also be considered an important element of the Czech economy. The number of people working in this sector in the last ten to fifteen years ranged between 400 and 470 thousand, which represented a share of 7.5 to almost 10 percent of the total number of employed persons in the Czech Republic [9]. Even from this point of view, however, it is true that in recent years the share of employed people in the construction industry has been slightly decreasing, when it peaked around 2010.

Construction is an industry that shows high sensitivity to the business cycle. Compared to the development of gross domestic product (Figure 1), it is very volatile. Year-on-year changes in construction output tend to exceed year-on-year changes in gross domestic product, in both directions. We were able to convince ourselves of this both after the financial and economic crisis of 2008 and 2009, and now, when the Czech economy is still coming to terms with the effects of the coronavirus crisis. A characteristic feature of the construction industry is also its lag behind the course of the economic cycle, especially in the phase when the economy is recovering from the crisis. Again, both data from the period after 2009 and current data testify to this. In addition to economic development, a number of other factors influence the construction industry. They affect both the demand and supply side of the construction market. If we ignore engineering construction, which includes, for example, transport and other infrastructure projects, then the demand for construction works and building materials is mainly influenced by the situation on the labour market, the development of household disposable income, interest rates for loan products for financing housing, regulation of the mortgage market, prices building plots, expectations of households and companies. The supply side is mainly influenced by regulation of the construction market (legislation regarding construction permits), availability of labour in the construction industry, production capacity of construction materials and the resulting price development of both construction materials and construction works. A major threat to the construction market is the freezing of the mortgage market that occurred last year. Although the mortgage market showed certain signs of recovery in the first quarter of 2023, it still lags significantly behind the development in the period before 2022.

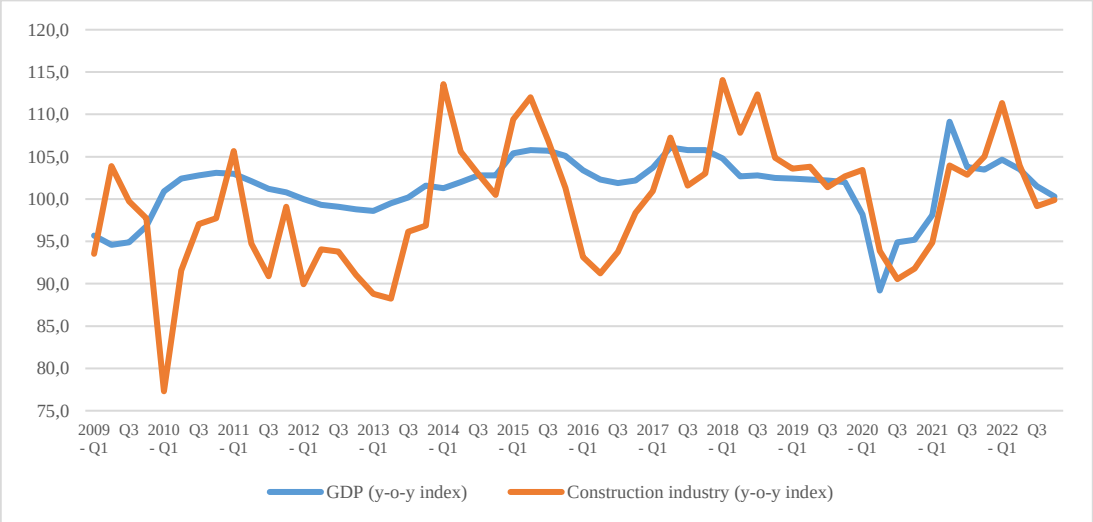


Figure 1 GDP and Construction Industry in the Czech Republic in the years 2009-2022.
Source: Czech Statistical Office [9], World Bank [17]

3 DEA Models and Construction Industry

In today's business environment, the evaluation and quantification of efficiency have become key factors in ensuring the effective operation of businesses. Through efficiency analysis, inefficient businesses can be examined in

detail and their inefficiencies addressed by identifying and eliminating their weaknesses. Farrell [10] introduced the concept of economic efficiency and distinguished it into two categories: technical efficiency and price (scale) efficiency. Achieving economic efficiency requires the simultaneous achievement of both technical and economies of scale. While efficiency is commonly assessed using various indicators, the prevailing approach often involves the use of data envelopment analysis (DEA) models.

DEA models, initially introduced by Charnes, Cooper, and Rhodes [3], build upon Farrell's concept, are extensively employed to evaluate the performance of diverse entities such as countries, regions, companies, schools, hospitals, insurance companies, and military units [5]. DEA models calculate the efficiency of decision-making units (DMUs) and generate two subsets: efficient and inefficient DMUs based on their efficiency scores. The key principle behind DEA models is to estimate the efficient frontier, which represents the optimal relative ratios of inputs and outputs for the compared DMUs. A DMU is deemed efficient if it resides on the efficient frontier, while inefficient DMUs fall below it and can enhance their efficiency by adjusting inputs or outputs [6].

Let us denote $\boldsymbol{Y} = \left(y_{rj}, r = 1, \dots, R, j = 1, \dots n\right)$ a non-negative matrix of outputs and $\boldsymbol{X} = \left(x_{sj}, s = 1, \dots, S, j = 1, \dots n\right)$ a non-negative matrix of inputs. The efficiency score of the unit j_0 is calculated using this model:

$$\begin{aligned} \text{Maximise} \quad & U_{j_0} = \frac{\sum_{r=1}^{R} u_r y_{r,j_0}}{\sum_{s=1}^{S} v_s x_{s,j_0}} \\ \text{subject to} \quad & \frac{\sum_{r=1}^{R} u_r y_{r,j}}{\sum_{s=1}^{S} v_s x_{s,j}} \leq 1, \quad j = 1, \dots, n, \\ & u_r \geq \varepsilon, \qquad r = 1, \dots, R, \\ & v_s \geq \varepsilon, \qquad s = 1, \dots, S, \end{aligned} \tag{1}$$

where u_r is a positive weight of the r-th output, v_s is a positive weight of the s-th input, and ε is an infinitesimal constant, U_{j_0} is called the efficiency score for j_0-th unit under evaluation. The U_{j_0} equals 1 for the efficient units and lower than 1 for the inefficient units. In model (1), the objective function is non-linear, which necessitates the linearization of the model. Typically, two approaches are employed for linearization based on the orientation of the linear model: input-oriented and output-oriented. This article, like similar papers, focuses on input-oriented models. The linearized version of the input-oriented model, commonly referred to as the CCR model [3], is presented as follows:

$$\begin{aligned} \text{Maximise} \quad & U_{j_0} = \sum_{r=1}^{R} u_r y_{r,j_0} \\ \text{subject to} \quad & \sum_{s=1}^{S} v_s x_{s,j_0} = 1, \\ & \sum_{r=1}^{R} u_r y_{r,j} - \sum_{s=1}^{S} v_s x_{s,j} \leq 0, \quad j = 1, \dots, n, \\ & u_r \geq \varepsilon, \quad r = 1, \dots, R, \\ & v_s \geq \varepsilon, \quad s = 1, \dots, S. \end{aligned} \tag{2}$$

The CCR input-oriented models are sometimes referred to as CRS as they assume constant returns to scale (CRS). It means that if the DMU with an input/output combination $(\boldsymbol{x}, \boldsymbol{y})$ is efficient, then the unit with an input/output combination $(\alpha\boldsymbol{x}, \alpha\boldsymbol{y})$, where $\alpha > 0$ is also efficient. Another possibility of returns to scale are the variable returns to scale (VRS), which were firstly mentioned in Banker et al. [1]. The DEA models with the variable returns to scale are called VRS or BCC models, according to Banker, Charnes and Cooper. The multiplicative form of the BCC input-oriented model (3) uses μ as a free variable and $U_{j_0}^V$ as the efficiency score for j_0-th unit under evaluation. The $U_{j_0}^V$ equals one for the efficient units. The inefficient units have an efficiency score lower than 1.

$$\begin{aligned} \text{Maximise} \quad & U_{j_0}^V = \sum_{r=1}^{R} u_r y_{r,j_0} + \mu \\ \text{subject to} \quad & \sum_{s=1}^{S} v_s x_{s,j_0} = 1, \\ & \sum_{r=1}^{R} u_r y_{r,j} - \sum_{s=1}^{S} v_s x_{s,j} + \mu \le 0, \quad j = 1,\ldots,n, \\ & u_r \ge \varepsilon, \quad r = 1,\ldots,R, \\ & v_s \ge \varepsilon, \quad s = 1,\ldots,S, \end{aligned} \tag{3}$$

The construction sector is not one of the most typical sectors for the use of DEA models, but there are nevertheless analyses related to efficiency in the construction sector, mainly in China. Xue at al. [18] combined DEA and the Malmquist productivity index to measure the productivity changes of Chinese construction industry from 1997 to 2003. Chen et al. [7] analyzed the energy efficiency in 30 Chinese provinces in the years 2003-2011 using three stage DEA models. Huo et al. [12] combined DEA models with the TFEE (total factor energy efficiency) model to measure the energy efficiency in the construction industry in the years 2006-2015 in 30 Chinese provinces. In addition to energy efficiency in the construction sector, DEA models were also used in country comparisons – for example Zhu et al. [19] provided a feasible pathway towards applying DEA in a multi-regional input-output analysis. They ranked 41 countries in the years 2000 and 2014 through measuring their overall economic performance. In addition to these global studies, the researchers also focused on comparisons of construction companies. Horta, Camanho and Costa [11] combined DEA models with KPIs benchmark scores on a sample of 20 Portuguese leading contractors. A two-stage analysis has been carried out by Tsolas [16], integrating the DEA framework with ratio analysis for modelling the profitability efficiency and effectiveness, respectively, of a sample of Greek-listed construction firms.

Based on a review of existing studies ([2], [3], [6]), usually three types of efficiencies could be used depending on the inputs and outputs: production, cost and profit efficiencies. On the basis of the available data, we have decided for the traditional approach, i.e. for the analysis of production efficiency, where the inputs usually include the number of employees, materials, inventories, liabilities, etc., while the outputs are usually sales, turnover or number of units produced.

4 Models and Data

We used two different data sources for the analysis. The first was data from the Albertina CZ Gold Edition database (Bisnode MagnusWeb, 2021), from which we obtained data for 2018 for firms belonging to NACE categories 2332 and 2361. As mentioned above, there are usually fewer entities in the 2332 category than in the 2361 category. In our case, there were 10 firms in 2332 – Manufacture of bricks, tiles and construction products, in baked clay (Table 1) and 57 firms in 2361 – Manufacture of concrete products for construction purpose (Table 2) for which all available data for efficiency (productivity) analysis were provided. The second data source was based on our previous research in comparison with CEEC Research company [14].

NACE 2332	**avg.**	**min.**	**max.**
Number of companies	10	x	x
I1: Number of employees	84.2	3	311
I2: Total liabilities (thous.CZK)	360757.7	118	1742978
I3: Inventory (thous.CZK)	51811.9	1	246391
O1: Revenue from sales of own products and services (thous. CZK)	311577.5	154	1752151

Table 1 Data for the NACE 2332 companies, year 2018

As there is not a lot of companies, we used 3 inputs and 1 output in Model 1 and 1 input and 3 outputs in Model 2 (see Table 3). It is not typical to have only 1 input or output in DEA models but usually, as in our case, it is influenced by the data availability – data must be easily obtained across all DMUs. The other reason to choose only 1 input / output is the fact that we consider economic indicators only and we would like to compare our results when we continue with the research using several financial analysis measurements (ratio analysis). Finally, the aim was to assess whether choosing a different number of inputs and outputs in the models and with data from two different sources but from the same sector would significantly affect the results or whether the results would be

similar regardless of the data. It is important to highlight that the data are economic in nature and refer to similarly oriented enterprises in the field of walling materials production, i.e. the homogeneity of DMUs is respected here.

The aim of this section was to assess whether firms were close to the efficiency frontier in this year, an exceptional year from a construction industry perspective, and whether the efficient firms included large manufacturers. To be comparable with Model 2, only companies with at least 1 employee were selected. In both cases, small, medium and large companies are represented. The largest companies (in terms of number of employees) are BEST, a.s., HELUZ cihlářský průmysl, v. o. s., PORFIX CZ a.s., Wienerberger, s. r. o., Xella CZ, s.r.o. and ŽPSV s.r.o.

NACE 2361	**avg.**	**min.**	**max.**
Number of companies	57	x	x
I1: Number of employees	74.2	1	532
I2: Total liabilities (thous.CZK)	221842	325	1687393
I3: Inventory (thous.CZK)	36336.6	0.5	256309
O1: Revenue from sales of own products and services (thous. CZK)	221988.8	1054	1704481

Table 2 Data for the NACE 2361 companies, year 2018

Since we were also able to obtain data on production and sales volumes in 2018 for the 21 most important manufacturers of masonry (walling) materials in cooperation with CEEC Research, in the second part of the analysis (Model 2) we decided to use this data (Table 4) for the analysis of production efficiency with the inclusion of these inputs (Table 3 - Model 2). The results of the two models are then compared.

	Inputs	**Outputs**
Model 1	I1: Number of employees I2: Total liabilities I3: Inventory	O1: Revenue from sales of own products and services (thousands of CZK)
Model 2	I1: Number of employees	O2: Net sales in thousands of CZK O3: Production volume (in m^3) O4: Annual sales in mil. CZK

Table 3 List of inputs and outputs for DEA models

Data Model 2	**avg.**	**min.**	**max.**
Number of companies	21	x	x
I1: Number of employees	173	4	744
O2: Net sales in thousands of CZK	589914	1789	2350607
O3: Production volume (in m^3)	101445	700	590000
O4: Annual sales in mil. CZK	304	2	1844

Table 4 Data for the Model 2

5 Results

In Model 1, we used only one output for all DEA models, namely revenue maximization, while the inputs were number of employees, liabilities and inventories. Efficiency measures were first calculated for the 10 firms falling in NACE area 2332, then for the 57 firms in NACE area 2361, and then for all 67 firms combined.

Among the 10 companies in the 2332 category, Wienerberger, s.r.o. is one of the largest, and was rated among the CCR and BCC efficient. However, the other two efficient firms are among those with less than 50 employees, i.e., given the size of the firms, larger firms were not found to be more efficient. The second largest company, HELUZ cihlářský průmysl, v.o.s., only achieved about 65% CCR and BCC efficiency. Due to the smaller number of firms, BCC efficiency is high, however, even by CCR it is evident that only 2 small firms achieved very low efficiency (lower than 20%), thus the average efficiency of all in the category 2332 - Manufacture of bricks, tiles and construction products, in baked clay - is quite high compared to the 2361 category (Table 5).

In category 2361, the competition is somewhat greater and none of the large firms with Manufacture of concrete products for construction purposes as their main activity were rated as efficient in CCR and BCC models. Only PORFIX CZ a.s., and Xella CZ, s.r.o. were rated as BCC efficient. Again, most of the BCC efficient firms are

small firms with less than 50 employees. However, the efficiency of the other two large firms (BEST, a.s., ŽPSV s.r.o.) was quite good at over 70%. However, it can also be noted that very low efficiency rates (less than 20%) were achieved by firms with less than 50 employees.

	Model 1:NACE 2332	**Model 1:NACE 2361**	**Model 1: all**	**Model 2**
Number of companies	10	57	67	21
Number of efficient (CCR)	3	5	5	2
Number of efficient (BCC)	6	11	14	5
Avg, CCR efficiency	64.87%	38.10%	36.42%	35.41%
Avg. BCC efficiency	91.08%	55.96%	55.91%	55.33%

Table 5 Results for the Model 1

Given the small size of group 2332 and the fact that the firms listed are typically engaged in both the manufacture of bricks (2332) and the manufacture of concrete products for construction (2361), we have combined the two categories together. The results remained similar, i.e. 3 large firms and a few small firms with less than 20 employees were classified as BCC efficient, with the efficiency rate for large firms being higher than for medium and small firms (with small firms marked as efficient with a few exceptions).

In Model 2, we used slightly different data to analyse production efficiency, for 21 major manufacturers according to CEEC Research. On the basis of 1 input (number of employees) and 3 outputs (net sales, annual production volume, annual sales volume), 2 firms were evaluated as CCR efficient (among them Xella CZ s.r.o.) and 5 as BCC efficient (among them Xella CZ, s.r.o., Wienerberger, s.r.o. and HELUZ cihlářský průmysl, v. o. s.,). Also in this analysis, 2 firms with 4 and 5 employees were among the efficient ones, while among the firms with low efficiency, mostly those with more than 10 and less than 50 employees appeared.

6 Conclusions

The construction industry plays a crucial and multifaceted role in society and the economy. Infrastructure development, job creation, economic growth, urban development and housing are key reasons why the construction industry is important. In this paper we focused on the largest producers of building materials in the Czech Republic and the development of the efficiency of these firms in the years 2018 using DEA models.

The aim is to assess the efficiency of companies in 2018, including whether large and established companies performed better than small and medium-sized ones. We focused primarily on brick and concrete building materials manufacturers, excluding entities with no employees from the analysis. Thus, only firms with more than 1 employee were included. To assess their efficiency, 2 datasets and DEA models were used. Although each model used different economic indicators, we concluded, that large firms use their resources more efficiently than small and medium-sized firms, with minor exceptions mainly including micro-enterprises with less than 10 employees.

Given the financial requirements of production, this result could be expected. The situation is thus different from, for example, the services sector, where, for comparison of the efficiency of accommodation and catering enterprises, small firms were more efficient [13]. Anyway, the above result was confirmed in both models, i.e. it could be shown that despite different data sets (on firms from the same sector) and the use of different inputs and outputs (different economic indicators) it did not have a significant impact on the resulting assessment of efficiency differences between large firms on the one hand and small and medium-sized firms on the other.

The aim of further research will be to assess the consistency of the results of DEA models with models used in financial analysis, in particular selected bankruptcy models, Altman's Z-score, IN95 and IN05. Already in previous research from 2010 and 2014, focusing on the construction sector, the authors have shown that large firms perform slightly better on these indicators than small and medium-sized firms [4]. This was also the reason why we focused on the differences between small, medium and large firms.

This research has its limitations, of course. None of the datasets included all firms in a given sector of building materials manufacturers, so only firms for which data were available were compared. Another limitation is that the classification in a given sector is related to the main activity of the firm, but given the similarities of some NACE areas, it is possible that some firms report a different area as their main business activity. Finally, the efficiency coefficients are affected by the choice of inputs and outputs in the DEA models and it was not always possible to use the most appropriate indicators or the same indicators as used in other research (due to lack of data).

Nevertheless, we consider this research as the beginning of further analyses of the efficiency of selected sectors and tracking the differences between small, medium and large firms.

Acknowledgements

The paper was supported by the contribution of long-term institutional support of research activities by the College of Polytechnics Jihlava.

References

[1] 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.
[2] 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, 91–107.
[3] Charnes, A., Cooper, W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2 (6), 429-444.
[4] Činčalová, S. & Jánsky, J. (2020). Evaluation of Finantial Health of Czech Construction Companies Using Prediction Models. In *63rd International Scientific Conference on Economic and Social Development Development: Book of Proceedings* (111-119), Zagreb: Croatioan Chamber of Economy.
[5] Cook, W. D. & Seiford, L. M. (2009). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research*, 192, 1–17.
[6] Cooper, W. W., Lawrence, M.S. and Zhu, J. (2004). *Handbook on Data Envelopment Analysis*. Norwell: Kluwer Academic Publishers
[7] Chen, Y., Liu, B., Shen, Y. & Wang, X. (2016). The energy efficiency of China's regional construction industry based on the three-stage DEA model and the DEA-DA model. *KSCE Journal of Civil Engineering*, 20, 34-47.
[8] Czech Statistical Office (2022). *Construction – time series.* [Online]. Available at: https://www.czso.cz/csu/czso/sta_ts [cited 2023-05-12]
[9] Czech Statistical Office (2023). *Statistical metainformation system.* [Online]. Available at: https://apl.czso.cz/iSMS/en/cisdet.jsp?kodcis=5105 [cited 2023-05-12]
[10] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120 (3), 253-290.
[11] Horta, I. M., Camanho, A. S. & Da Costa, J. M. (2010). Performance assessment of construction companies integrating key performance indicators and data envelopment analysis. *Journal of Construction engineering and Management*, 136(5), 581-594.
[12] Huo, T., Tang, M., Cai, W., Ren, H., Liu, B. & Hu, X. (2020). Provincial total-factor energy efficiency considering floor space under construction: An empirical analysis of China's construction industry. *Journal of cleaner production*, 244, 118749.
[13] Kuncová, M., Zýková, P., Kozáková, P. & Lízalová, L. (2022). Analysis of the efficiency of the Czech companies in the NACE sector "Accommodation and food service activities". *In 40th International Conference Mathematical Methods in Economics 2022-Proceedings* (205-211). Jihlava: College of Polytechnics.
[14] Musil, P., Kuncová, M., Činčalová, S., Rojík, S. & Dostál, J. (2021). Analýza stavebního trhu se zaměřením na oblast zdicích materiálů v České republice a Slovenské republice (Analysis of the construction market with a focus on walling materials in the Czech Republic and Slovak Republic). *Research report.*
[15] Stavbaweb.cz (2018). *Vývoj na trhu cihelných výrobků v České republice v roce 2018 (Development of the brick products market in the Czech Republic in 2018*). [Online]. Available at: Stavbaweb.cz – Vývoj na trhu cihelných výrobků v České republice v roce 2018 [cited 2023-05-12]
[16] Tsolas, I. E. (2011). Modelling profitability and effectiveness of Greek-listed construction firms: an integrated DEA and ratio analysis. *Construction Management and Economics*, 29 (8), 795-807.
[17] World Bank (2023). *GDP growth (annual %) – Czechia*. [Online]. Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=CZ [cited 2023-05-12]
[18] Xue, X., Shen, Q., Wang, Y. & Lu, J. (2008). Measuring the productivity of the construction industry in China by using DEA-based Malmquist productivity indices. *Journal of Construction engineering and Management*, 134 (1), 64-71.
[19] Zhu, R., Hu, X., Li, V. & Liu, C. (2021) Investigating economic roles of multinational construction industries: A super-efficiency DEA approach, *Applied Economics*, 53 (41), 4810-4822

Self-learning Metaheuristics for Pareto Front Approximation

Marek Kvet[1], Jaroslav Janáček[2]

Abstract. This paper reports recent research and development in the field of specific location problems, in which two contradictory objectives need to be minimized. Considering two conflicting criteria in the mathematical model leads to the necessity of constructing a special subset of feasible solutions called a Pareto front. Each pair of its elements must hold the non-dominance property. Since obtaining the complete Pareto front proved to be computationally difficult, the experts' attention has been paid to the development of various approximate approaches including metaheuristics and hyperheuristics. The content of this paper focuses on a family of self-learning metaheuristics based on minimization of an area determined by a set of non-dominated solutions for approximation of the Pareto front of bi-criteria location problems.

Keywords: Discrete location problems, conflicting criteria, Pareto front approximation methods, metaheuristics

JEL Classification: C44, C61
AMS Classification: 90C05, 90C06, 90C10, 90C27

1 Introduction

Pareto front of designs of two-criterion public service systems plays an important role in management of public affairs, because it enables a responsible decision-maker a qualified negotiation with public representatives to balance the points of view of the public majority and the worst situated minority [1, 3, 5]. Completion of the Pareto front using exact methods is complex and computational time demanding task. The usage of heuristics is a suitable option, especially in the case when the two-criterion problem belongs to the family of combinatorial problems. Efficiency of sophisticated heuristics for obtaining a good Pareto front approximation mostly depends on the heuristic parameter settings. Finding of proper parameter values can be done by series of previous experiments with the heuristic applied on similar problems or it can be performed by a self-learning process applied during the problem solution.
The scheme of the used incremental process is based on principle of gradual refinement [10], which step-by-step improves the current Pareto front approximation by including a newly obtained non-dominated problem solution. A simple step of the process consists of processing one member of the approximation, what means that the member is used as a starting solution of a routine, which performs series of swap operations.

This paper is devoted to study and comparison of three metaheuristics, which apply a self-learning process to set up their parameters. The self-learning process is characterized by an entity called memory, in which experience obtained in history of the process is accumulated and used for control of the next steps. The memory may have various forms and can be updated in various ways. The different memory definitions and usage will be studied from different points of view.

2 Gradual Refinement Scheme for Obtaining a Good Approximation of Pareto Front

Numerous Pareto front approximation methods proposed adhere to the principle of continuous updates of the resulting collection of non-dominated system designs, also known as *NDSS* (non-dominated solutions set). Let us talk about some specifics. Consider two contradictory objectives f_1 and f_2. The updating procedure begins with an initial non-empty collection of non-dominated solutions that are arranged in ascending value of f_2 order. The candidate solutions C characterized by $(f_1(C), f_2(C))$ is tested to find its position between predecessor P and successor S in the current set of non-dominated solutions so that $f_2(P) < f_2(C) \leq f_2(S)$. If $f_1(P) \leq f_1(C)$, then the candidate C is

[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

dominated by predecessor *P* and is abandoned. Otherwise, if the candidate *C* is not dominated by the successor *S*, the candidate joins the updated set as a new member. The following members of the sequence starting with the successor *S* are compared to *C* and if *C* dominates them, they are excluded from the updated set.

The idea of the region produced by specific *NDSS* set pieces may be applied if the Pareto front approximation's quality is to be assessed [8, 10, 11, 14]. It goes without saying that in order to make the evaluation accurate and meaningful, the bordering elements of the set must be determined correctly. More information is covered in [9].

As far as concrete forms of the mentioned objective functions f_1 and f_2 are concerned, they follow the fundamental guidelines for Emergency Medical Service (EMS) system designing used also in our previous research activities, the results of which are published in [6, 7, 8, 9, 10, 11, 14] and in numerous other.

The gradual refinement procedure begins with a small list of non-dominated solutions, which may consist of only two elements. The current collection *NDSS* of *noNDSS* non-dominated solutions is initialized by the starting set and is continuously updated by a simple sequential process. The basic procedure processes current *NDSS* from the solution $\mathbf{y}^1$ to the solution $\mathbf{y}^{noNDSS-1}$ and continuously updates the current set of non-dominated solutions. In one step, solution $\mathbf{y}^k$ temporarily located at the *k*-th position of *NDSS* is used as an initial solution for a local search application, which produces candidates for *NDSS* updating. After the local search has finished, the *k*-th solution of *NDSS* may be changed. In such a case, the local search continues with the changed solution, otherwise the (*k*+1)-th solution is processed. The basic procedure terminates, when *k*+1 = *noNDSS*. The basic procedure can be repeated with the resulting *NDSS* until a given computational time limit is exceeded.

As the Pareto front completion is a hard computational problem, we concentrate our effort on establishing a good approximation of it. The approximating collection of non-dominated solutions (*NDSS*) will be represented by a sequence of *noNDSS* solutions $\mathbf{y}^1$, …, $\mathbf{y}^{noNDSS}$ ordered according to increasing values of f_2. Here, the symbol *noNDSS* represents the cardinality of the *NDSS* set and it is assumed to be non-negative integer. To obtain a relevant approximation, the bordering solutions $\mathbf{y}^1$ and $\mathbf{y}^{noNDSS}$ must be determined to be very close to the most left and the most right solutions of the Pareto front as concerns the values of f_1 and f_2. Under these assumptions, the quality of the approximation *NDSS* can be measured by *A*(*NDSS*) computed according to the expression (1).

$$A(NDSS) = \sum_{k=1}^{noNDSS-1} \left(f_1\left(\mathbf{y}^k\right) - f_1\left(\mathbf{y}^{noNDSS}\right)\right)\left(f_2\left(\mathbf{y}^{k+1}\right) - f_2\left(\mathbf{y}^k\right)\right) \tag{1}$$

The same formula can be used to compute the area *A*(*PF*) of the complete original Pareto front (*PF*) and it holds that the *A*(*PF*) is a lower bound of any *A*(*NDSS*). The absolute value of area can be used as a metric to compare various *NDSS* sets consisting of different number of their elements [9].

To make the formal description of suggested algorithms complete, let us define *u*(*NDSS*, $\boldsymbol{x}$) as such new *NDSS*, which results from the updating process. If the candidate solution $\boldsymbol{x}$ is dominated by at least one *NDSS* member, current set of non-dominated solutions stays unchanged. In the opposite case, the solution $\boldsymbol{x}$ becomes a new element of NDSS and all elements dominated by $\boldsymbol{x}$ are excluded from the set [8, 9, 10, 11, 14].

Considering the order of *NDSS* solutions, one can easily implement an algorithm, which decides in *noNDSS* steps whether an arbitrary feasible solution $\boldsymbol{x}$ is dominated by an element of the current *NDSS* or it can be included into the *NDSS* improving the associated area denoted by *A*(*NDSS*). Such algorithm can be used for iterative sequential updating the initial *NDSS* under the condition of having a source of candidate solutions.
The suggested decrementing algorithm is based on a neighborhood search, when the inspected neighborhood *N*($\boldsymbol{y}$) of a current solution $\boldsymbol{y}$ consists of all feasible solutions, which can be obtained by performing a simple permitted operation with the current solution $\boldsymbol{y}$.

3 Self-Learning Routines

3.1 Routine using simulated annealing elements

The routine RSA starts with the processed solution $\boldsymbol{y}$, which is a member of the input *NDSS*. The next series of control parameters is considered: Current state *s* of the learning process, γ - size of possible change of the state in one step, α - parameter of forgetting, β - parameter of learning intensity, *Thr* - parameter used in acceptance rule of the routine.

At the beginning of the gradual refinement scheme, it is assumed that the initial *NDSS* consists only of two members, which correspond with the most left and the most right Pareto front members. The initial value of *s* is zero

and the parameters *Thr*, α, β and γ are set at starting values. The routine returns updated *NDSS*, *Thr* and *s*. more details are discussed in [8]

RSA(***y***, *NDSS*, *Thr*, *s*, α, β, γ)

0. Set $A0 = A(NDSS)$, $Thr0 = Thr$ and perform random change of *Thr* according to the result of the random trial with probability *prob*, which is determined by the rule: $prob = 0$ for $s < -1$, $prob = 1$ for $s > 1$ and $prob = (1+s)/2$ for $s \in [-1, 1]$. If randomly generated value *R* from the interval [0, 1] satisfies $R < prob$, then $Thr = Thr0+\gamma$, otherwise $Thr = Thr0 - \gamma$.
1. Perform the neighborhood search algorithm with the starting solution ***y***, where a neighbor of the current solution is an arbitrary solution, which differs from the current one in one element. Each neighbor ***x*** is tested as a potential improving solution for *NDSS* updating. The move from the current solution to the neighbor ***x*** is performed whenever randomly generated *R* from interval [0, 1] satisfies the acceptation rule $R \leq exp(A(NDSS)-A(u(NDSS, \boldsymbol{x})) - Thr)$.
2. Set $A1 = A(NDSS)$ and update *s* according to the formula $s = \alpha s + \beta sign((A0 - A1 - Thr)(Thr - Thr0))$

3.2 Routine based on ant colony optimization

A classical ant colony optimization algorithm simulates performance of ant colony members in searching the shortest path to food located in nodes of a transportation network. An individual ant chooses its way in the network randomly, but it takes into consideration heuristic information about advantage of its move to the next node and experience of the previous ants, which is expressed by pheromone layer connected to the potential move. At the end of the ant search, the ant lays updates pheromone layer on the inspected path. An evaporating process updates the pheromone layer. The core of ant colony application consists in mapping the solved problem on a network [11].

One run of the routine ACO carries out work of one ant in the network, where starting node of the ant is given by the processed solution ***y*** and by searching strategy *t*, determined by a combination of two parameters *Thr*(*t*) and *maxNos*(*t*). Parameters *Thr* gives minimal improvement to consider a move admissible and the parameter *maxNos*, gives the number of admissible moves, from which the best one is realized. The next nodes of the ant's network are represented by current solutions obtained by the swap operation, which replaces one location of a current solution by a location, which is not included in the solution. Instead of recording the inspected path in detail, the path description is reduced to the set of locations, which have been subsequently included into the starting solution. These recorded entries together with chosen strategy are considered in the phase of laying pheromone.

The routine ACO starts with the processed solution ***y***, which is a member of input *NDSS*. The following control parameters are considered: *S* - pheromone layers of strategies 1, …, *noS* at disposal, *S*(*t*) denotes pheromone layer of the strategy *t*. *F* - pheromone layers of locations used in moves from one to another current solution, *F*(*j*) denotes pheromone layer of the location *j*. ρ - intensity of evaporation, which gets values from interval [0, 1]. τ - scaling parameter for pheromone laying.

At the beginning of the gradual refinement scheme, it is assumed that the initial *NDSS* consists only of two members, which correspond with the most left and most right Pareto front members. The initial value of pheromone equals to one and the parameter ρ is set at starting value. The routine returns updated structures *NDSS*, *S* and *F*. A detailed description of the ant's search and pheromone laying follows.

ACO(***y***, *NDSS*, *S, F,* ρ, τ)

0. Set $A0 = A(NDSS)$, choose strategy *t* from the set of strategies at disposal using roulette wheel random trial, where the probability of strategy *t* is $S(t)/\sum_{k=1}^{noS} S(k)$.
1. Perform the neighborhood search algorithm with starting solution ***y*** and strategy given by *Thr*(*t*) and *maxNos*(*t*). A neighbor of the current solution is an arbitrary solution, which differs from the current one in one element. Each neighbor ***x*** is tested as a potential improving solution for *NDSS* updating. The move from the current solution to the neighbor ***x*** is performed in accordance to a random trial with probability *CFit*/(*CFit* +*MFit*) in favor of the candidate. *MFit* and *CFit* are fitness values of the recently appointed swap operation and the tested one. The fitness value *CFit* for the swap operation inserting location *j* in a current solution and resulting in candidate ***x*** is computed according to the formula $CFit = Dec*F(j)$, where $Dec = A(NDSS)-A(u(NDSS, \boldsymbol{x}))$. The ant's search finishes, when either whole neighborhood is inspected or *maxNos*(*t*) candidates are evaluated.

2. Set $A1 = A(NDSS)$ and define $Dec = A0 - A1$ and update the pheromone layer $S(t)$ of the chosen strategy t and $F(j)$ of all inserted locations j using the following formulae: $S(t) = S(t) + Dec/\tau$, $F(j) = F(j) + Dec/\tau$. The final pheromone adjustment is performed with pheromone layer of each object according to the formula $S(t) = (1 - \rho)*S(t)$ and $F(j) = (1 - \rho)*F(j)$, where ρ is a evaporating coefficient.

3.3 Selective hyperheuristic routine

The suggested selective hyperheuristic [14] disposes with a set R of search algorithms. An algorithm $r \in R$ has its own score denoted $Score(r)$. The score is initialized by a small positive value at the beginning of the hyperheuristic and also τ is determined as $A(NDSS)$, where $NDSS$ consists only of two members. The neighborhood search algorithm r performs according to the strategy given by $Thr(r)$ and $maxNos(r)$. Let $\boldsymbol{y}$ denote the processed solution $\boldsymbol{y}$, which is a member of the input $NDSS$. The following steps describe the core of the suggested selective hyperheuristic.

$SHR(\boldsymbol{y}, NDSS, Score, \tau)$

0. Set $A0 = A(NDSS)$, choose search algorithm r from the set R at disposal using roulette wheel random trial, where the probability of algorithm r is $Score(r)/\sum_{k=1}^{noS} Score(k)$.
1. Perform the neighborhood search algorithm with starting solution $\boldsymbol{y}$ and strategy given by $Thr(r)$ and $maxNos(r)$. A neighbor of the current solution is an arbitrary solution, which differs from the current one in one element. Each neighbor $\boldsymbol{x}$ is tested as a potential improving solution for $NDSS$ updating. The move from the current solution to the neighbor $\boldsymbol{x}$ is performed in accordance to highest value of $A(NDSS)$-$A(u(NDSS, \boldsymbol{x}))$ from the $maxNos(r)$ neighbors, for which the difference is greater than $Thr(r)$. The search finishes, when either whole neighborhood is inspected or $maxNos(r)$ candidates are evaluated.
2. Set $A1 = A(NDSS)$ and define $Dec = A0 - A1$ and update the $Score(r)$ by $Score(r) = Score(r) + Dec/\tau$.

4 Computational Study

4.1 Benchmarks, objective functions and solving tools

The experiments reported in this study were performed on a common PC equipped with the Intel® Core™ i7-3610QM CPU@2.30 GHz processor and 8 GB RAM. The algorithms were implemented in Java programming language in he NetBeans IDE 8.2 software.

Suggested routines have been verified on common benchmarks representing existing EMS system operated in eight higher territorial units (HTU) of Slovakia. The same instances were used in our previous research reported in [6, 7, 8, 9, 10, 11, 14]. The list of problem instances contains the HTU of Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA). In the used input data, all inhabited network nodes represent the set of possible EMS station candidate locations and the patients' locations as well. The associated service is provided through the road network. The sizes of the problem instances are reported in the left part of Table 1.

Let us focus on the used objective functions, now. To formulate them mathematically, let I denote a finite set of candidates from which exactly p elements are to be chosen. Each subset $P \subseteq I$, the cardinality of which equals p, represents a feasible solution of the problem. In this computational study, we deal with so-called system and fair criteria. The system objective function can be formulated according to (2), where J stands for a finite set of served users emerging b_j requests for service. Moreover, let t_{ij} correspond to the traversing time from an EMS station located at $i \in I$ to the patient's location j and finally, let q_k denote the probability value of the case that the k-th nearest station will be the first available one [12, 13]. The operator $\min_k\{V\}$ gives the k-th smallest value in V.

$$f_1(P) = \sum_{j \in J} b_j \sum_{k=1}^{r} q_k \min_k \left\{ t_{ij} : i \in P \right\} \tag{2}$$

The second criterion f_2 takes into account the aspect of fairness [2, 4]. It minimizes the number of users' calls, for which the response time exceeds given threshold T. This objective $f_2(P)$ can be expressed by the formula (3).

$$f_2(P) = \sum_{j \in J} b_j \max\left\{0,\ sign\left(\min\left\{t_{ij} : i \in P\right\} - T\right)\right\} \tag{3}$$

The system and fairness criteria are usually in a conflict, which means that the better value of one of the criteria is paid for by worsening the value of the other objective function. Therefore, the Pareto front is usually provided as a selection of different system designs with clear consequences of one objective function value improvement on the other characteristic [6, 7, 12]. As the objective (2) follows the concept of generalized disutility, the parameter r was set to 3. The associated coefficients q_k were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. Parameter T used in the fair criterion (3) was set to the value of 10 minutes [9, 10, 12].

The following Table 1 summarizes the basic characteristics of used benchmarks and it contains also the information about the exact Pareto fronts. For each studied instance corresponding to one row of the table, the first two columns denoted by $|I|$ and p contain the problem size. The middle part of Table 1 contains the number of solutions *NoS* forming the Pareto front *PF* and the associated value of $A(PF)$. Finally, the right part summarizes additional information about the bordering points of *PF*. Let the symbol *MLM* denote the most left member and let *MRM* denote the most right one. Then, the objective function value (2) for *MLM* will be denoted as $f_1(MLM)$ and the criterion (3) for *MLM* will be denoted by $f_2(MLM)$. We will use an analogical denotation also for *MRM*.

HTU	$\|I\|$	p	*NoS*	$A(PF)$	$f_1(MLM)$	$f_2(MLM)$	$f_1(MRM)$	$f_2(MRM)$
BA	87	14	34	569039	42912	0	26649	280
BB	515	36	229	1002681	53445	453	44751	935
KE	460	32	262	1295594	61241	276	45587	816
NR	350	27	106	736846	59415	557	48940	996
PO	664	32	271	956103	65944	711	56703	1282
TN	276	21	98	829155	45865	223	35274	567
TT	249	18	64	814351	48964	450	41338	921
ZA	315	29	97	407293	48025	254	42110	728

Table 1 Benchmarks characteristics and the exact Pareto fronts description

4.2 Results of numerical experiments – routines comparison

The routines explained in the theoretical part of this paper were studied also from the practical point of view. The main goal of performed numerical experiments was to study their efficiency and accuracy. In this subsection, we provide the readers with the most important results. Obviously, more details can be found in [8, 11, 14].

To make the algorithms comparable, each of them was restricted to five minutes of computation and then, the resulting *NDSS* sets were compared. The obtained results are summarized in Table 2. Each column corresponds to one studied routine. To make the comparison comfortable, instead of absolute values of particular $A(NDSS)$ we computed so-called gaps. *Gap* is generally defined in percentage and it can be evaluated by the expression (4). If any routine contained randomness, the algorithm was run ten times and the average results were taken into account.

$$gap = 100 * \frac{A(NDSS) - A(PF)}{A(PF)} \tag{4}$$

HTU	RSA routine	ACO routine	SHR routine
BA	1.94	1.47	1.47
BB	1.46	1.03	0.97
KE	2.1	2.21	2.86
NR	1.82	0.72	0.88
PO	1.29	2.30	2.88
TN	1.44	0.72	1.45
TT	0.31	0.07	0.37
ZA	0.29	0.30	0.17

Table 2 Results of numerical experiments – table of average gaps

The results reported in Table 2 show that all suggested routines represent suitable solving tools for Pareto front approximation and they are able to provide very accurate approximation within five minutes. This feature makes them very useful for practical usage.

5 Conclusions

The research reported in this paper was focused on public service system optimization with two contradictory objectives, which cannot be optimized simultaneously. For such problems, the Pareto front seems to be a sufficient basis for responsible authorities to find the resulting service center deployment. The main scientific goal of this paper was aimed at the Pareto front approximation methods. Since the set of non-dominated system designs can be constructed in many different ways, the attention was paid to the comparison of three different routines. Based on performed experimental case study, it has been found that all suggested algorithms are suitable for practical usage thanks to their efficiency and accuracy. Only a short computational time is necessary to obtain good approximation of the original Pareto front.

Future research in the field of bi-criteria optimization could be aimed at development of other approaches for good Pareto front approximation. Another research topic could focus on generalization of current approaches and their adjustment for multi-objective system designing.

Acknowledgements

This work was supported by the following grants: VEGA 1/0216/21 "Designing of emergency systems with conflicting criteria using tools of artificial intelligence", VEGA 1/0077/22 "Innovative prediction methods for optimization of public service systems", and VEGA 1/0654/22 "Cost-effective design of combined charging infrastructure and efficient operation of electric vehicles in public transport in sustainable cities and regions". This paper was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

References

[1] Arroyo, J. E. C., dos Santos, P. M., Soares, M. S. & Santos, A. G. (2010). A Multi-Objective Genetic Algorithm with Path Relinking for the p-Median Problem. In: *Proceedings of the 12th Ibero-American Conference on Advances in Artificial Intelligence*, 2010, pp. 70–79.

[2] Bertsimas, D., Farias, V. F. & Trichakis, N. (2011). The Price of Fairness. In *Operations Research*, 59, 2011, pp. 17–31.

[3] Brotcorne, L., Laporte, G. & Semet, F. (2003). Ambulance location and relocation models. *Eur. Journal of Oper.Research*, 147, pp. 451–463.

[4] Buzna, Ľ., Kohání, M. & Janáček, J. (2013). Proportionally Fairer Public Service Systems Design. In: *Communications - Scientific Letters of the University of Žilina* 15(1), pp. 14–18.

[5] Current, J., Daskin, M. & Schilling, D. (2002). Discrete network location models, Drezner Z. et al. (ed) *Facility location: Applications and theory*, Springer, pp. 81–118.

[6] Grygar, D. & Fabricius, R. (2019). An efficient adjustment of genetic algorithm for Pareto front determination. In: *TRANSCOM 2019: conference proceedings*, Amsterdam: Elsevier Science, pp. 1335–1342.

[7] Janáček, J. & Fabricius, R. (2021). Public service system design with conflicting criteria. In: *IEEE Access: practical innovations, open solutions*, ISSN 2169-3536, Vol. 9, pp. 130665–130679.

[8] Janáček, J. & Kvet, M. (2023). Adaptive Parameter Setting for Public Service System Design. In: *Strategic management and its support by information systems 2023*, in print

[9] Janáček, J. & Kvet, M. (2021). Quality measure of Pareto front approximation. In: *Quantitative methods in economics: Multiple criteria decision making 21: Proceedings of the international scientific conference*, Bratislava: Letra Edu, pp. 95–103.

[10] Janáček, J. & Kvet, M. (2022). Repeated refinement approach to Bi-objective p-location problems. In: INES 2022: *26th IEEE International Conference on Intelligent Engineering Systems*, New York: Institute of Electrical and Electronics Engineers, pp. 41–45.

[11] Janáček, J. & Kvet, M. (2023). Ant colony optimization of Pareto front approximation. In: *IFORS 2023: The 23rd Conference of the International Federation of Operational Research Societies*, 10.-14.7.2023, in print

[12] 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*, pp. 163–167.

[13] Kvet, M. (2014). Computational Study of Radial Approach to Public Service System Design with Generalized Utility. In: *Proceedings of International Conference DT 2014*, Žilina, Slovakia, pp. 198–208.

[14] Kvet, M. & Janáček, J. (2023). Hyperheuristic as Tuning Tool of Generalized Swap Strategy. In: *META 2023: The 9th International Conference on Metaheuristics and Nature Inspired Computing*, 1.-4.11. 2023, in print

Socioeconomic Determinants of Electric Vehicle Adoption in Czechia

Jindřich Lacko [1]

Abstract. In an effort to reduce transportation-related CO_2 emissions to levels below those of 1990, the European Union and the Czech Republic are prioritizing the transition from internal combustion engines to battery-operated and hybrid propulsion vehicles. This study examines Czech vehicle registration data in the retail sector from 2019 to 2022, utilizing a stepwise linear regression approach to develop a model that explains the rate of electric vehicle adoption based on 37 socioeconomic predictors across 206 Czech administrative regions. The resulting model identifies five significant predictors, demonstrates a robust goodness of fit, and effectively mitigates the spatial autocorrelation initially observed in the data.

Keywords: Battery electric vehicles, Hybrid vehicles, Technology adoption and diffusion, Czechia

JEL Classification: Q54, C31
AMS Classification: 62M30

1 Introduction

Reducing CO_2 emissions significantly in comparison to 1990 levels is a target of both global and regional initiatives. A key factor for achieving this aim is reducing the volume of CO_2 emissions from transportation. This is in line with the contribution transportation related emissions have made to the current stock of CO_2 in the Earth's atmosphere. In order to achieve these aims it will be necessary to significantly reduce the volume of active Internal Combustion Engine (ICE) vehicles on EU roads. While alternatives to personal vehicle ownership such as mass transit and bicycling are certain to be a part of the solution it is unlikely that the current level of vehicle ownership (over 6 million personal vehicles are registered in the Czech Republic alone, a country of 10 million inhabitants [18]) could be feasibly reduced to zero.

As a consequence any plan to reduce and / or eliminate ICE personal vehicle ownership and use must include introducing of alternatives, such as Electric Vehicles (EV's). A number of studies was performed to evaluate the effectiveness of incentives in reducing ICE and promoting EV vehicle ownership. These were mostly focused on United States [8], [9]. In the European context important data come from Norway (global leader in EV adoption) [10], [16]. In Central Europe a review was performed of Polish customer preferences [5], [19].

This article presents a study of adoption of EVs in the Czech Republic administrative areas, as explained by socioeconomic characteristic of these regions. Dynamics of EV adoption are estimated using linear regression. Out of total of 37 possible predictors a subset was selected using technique of stepwise regression.

2 Methodology

The data were collected from the Czech Ministry of Transport Open Data [13] monthly snapshots. For the purpose of our analysis only registrations from years 2019 to 2022 were used, since EV registrations in earlier periods were negligible. The vehicle registration dataset provides spatial breakdown to the level of municipalities with extended powers (ORP).

We selected only personal vehicles, defined as M1 category in the relevant regulation [1]. Vehicle registrations in this category were further split into retail segment, defined as vehicle registrations by individuals (for private purposes) and non-retail segment, defined as vehicle registrations by companies. The retail segment also includes personal vehicles operated by individuals under leasing contracts.

Detailed breakdown of EV over brands and models is presented in Table 1.

[1] Vysoká škola ekonomická v Praze / Katedra Ekonometrie, jindra.lacko@vse.cz

Brand	Model	Retail	Non-Retail
ŠKODA	ENYAQ 80	58	1 512
ŠKODA	ENYAQ RS	11	629
ŠKODA	ENYAQ 80X	30	332
ŠKODA	ENYAQ 60	6	202
ŠKODA	other models	0	1
TESLA	MODEL 3	232	692
TESLA	MODEL Y	48	206
TESLA	MODEL S	56	157
TESLA	MODEL X	18	115
TESLA	other models	19	33
TOYOTA	YARIS HYBRID	334	591
TOYOTA	PRIUS	198	63
TOYOTA	PRIUS PLUS	36	18
TOYOTA	other models	85	46
NISSAN	LEAF 40KWH	28	171
NISSAN	LEAF	47	28
NISSAN	other models	42	60
RENAULT	CAPTUR E-TECH PLUG-IN HYBRID	17	112
RENAULT	CLIO E-TECH HYBRID	14	101
RENAULT	MEGANE E-TECH PLUG-IN HYBRID	2	54
RENAULT	other models	8	19
BMW	IX XDRIVE40	2	71
BMW	IX3	3	70
BMW	other models	17	108
SEAT	SEAT LEON SP E-HYBRID150	0	190
SEAT	other models	0	34
other brands	other models	147	435
		1 458	6 050

Table 1 Model Breakdown of EV Registrations

In order to model consumer activity we focused on retail registrations. Electrification of corporate fleets is expected to contribute materially to the overall reduction of transport CO_2 emissions, but it is best understood as a separate problem, impractical to model based on local socioeconomic variables.

For a deeper illustration consider comparison of Retail vs. Non-Retail EV penetration (Figure 1), noting the high penetration of EV registrations in Mladá Boleslav and Vrchlabí regions; these can be explained better as location of facilities of Škoda Auto than as result of local socioeconomic factors.

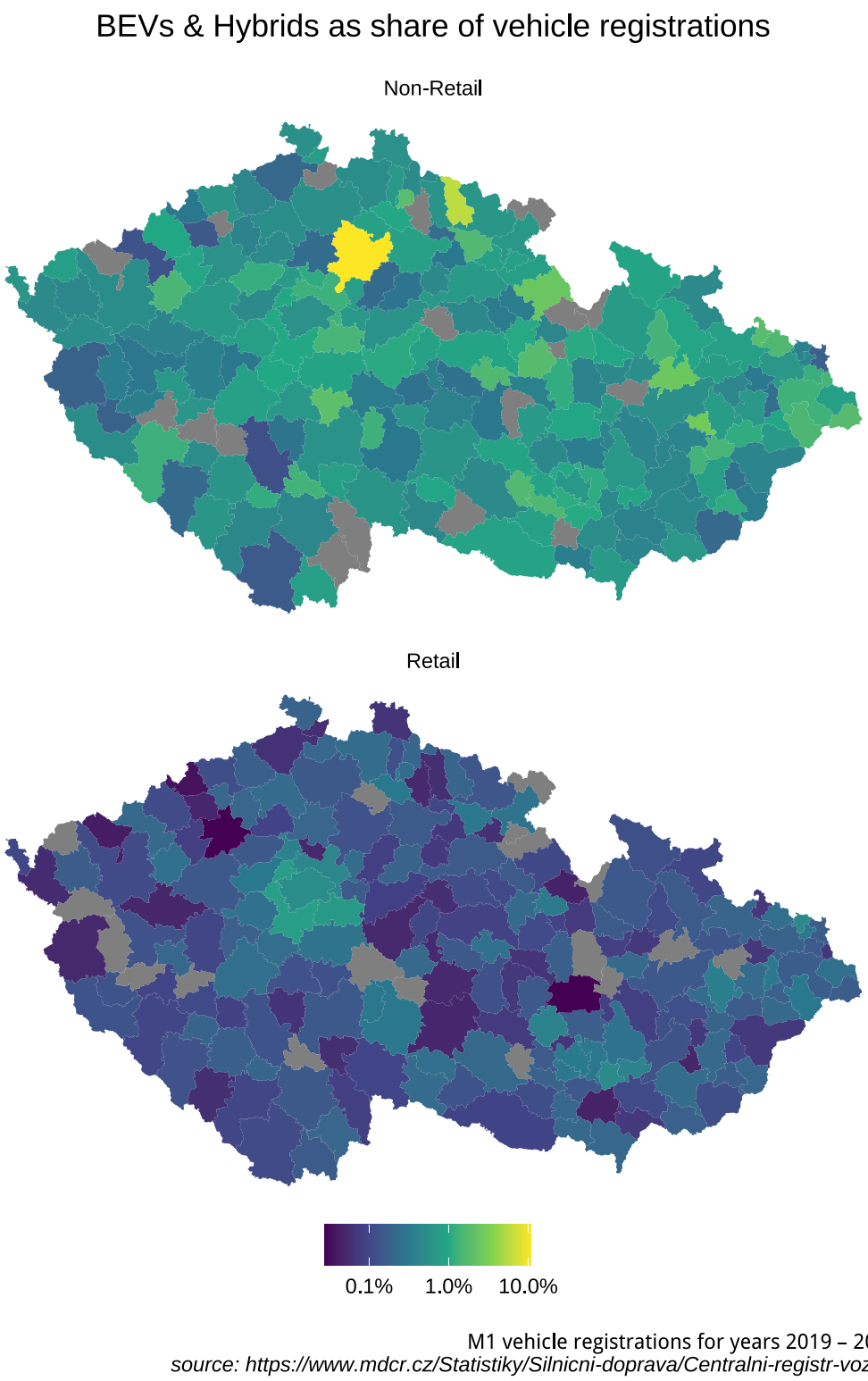


Figure 1 Non-retail vs. Retail EV Penetration (log scale) in Czech ORPs (n = 206)

For the socioeconomic predictors we used results of the previous (2011) census [6]; the results of the 2021 census were not available at the time of writing in the necessary level of detail. The census data were accessed from the Statistical Office API using *czso* package [4]. Data processing and modelling was done in statistical programming language R [15]. Modelling was performed using package *leaps* [12] for stepwise regression. Bayesian information criterion [17] was used as primary metric for variable selection.

Given that the population of the 206 Czech ORPs varies by several orders of magnitude (from more than a million

in the capital to less than 10 thousands in Pacov or Králíky regions) it was impractical to model registrations in terms of total registrations per region. Instead relative penetration was used – EV registration as percentage of total personal vehicle registration. Likewise socioeconomic predictors were normalized by population of the ORP.

A basic analysis of spatial dependence – which, if confirmed, would violate two key assumptions of ordinary least squares method, namely homoskedasticity and spatial independence of residuals – was performed using package *RCzechia* [11]. Spatial error was diagnosed via Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity [2]. Spatial autocorrelation of EV penetration at ORP level was measured using Moran's I statistic (1), as defined in [3].

$$I = \frac{n \times \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \times \sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

where $x_i, i = 1, \ldots, n$ are the n observations, w_{ij} are the spatial weights and S_0 is the sum of spatial weights $\sum_{i,j=1}^{n} w_{ij}$.

The value of the Moran's I statistic of Retail EV penetration was calculated via package *sfdep* [14], using queen contiguity neighborhood definition for determining the adjacency matrix of individual ORPs in order to evaluate spatial dependency of observed data.

3 Result

During the period in question total of 1 451 075 personal vehicles were registered, of which 7 508 could be considered EV's (5 004 battery operated EVs and 2 504 hybrid EVs).

Majority of the EV registrations come from the non-retail segment: 6 050 registrations. Compared to this the retail segment had only 1 458 registrations, which is 19.42% of total EV registrations over the period in consideration.

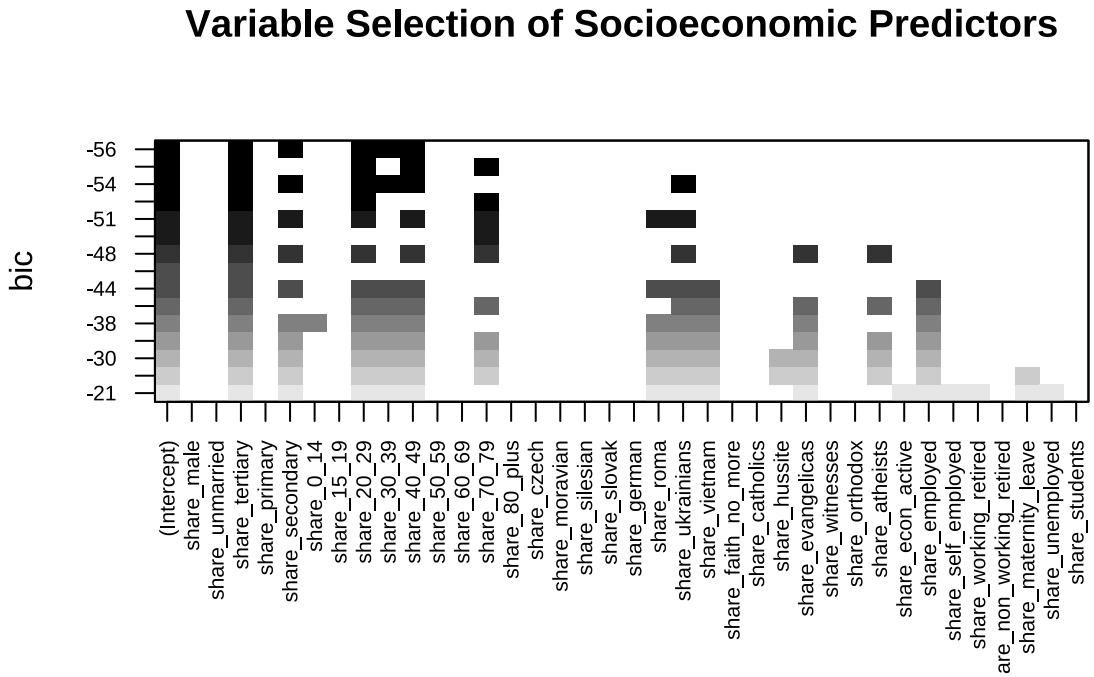


Figure 2 Schwartz's information criterion / stepwise regression

Out of the total 37 socioeconomic variables under consideration we identified the lowest BIC (-56.30301) for model containing intercept and 5 predictors. The predictors and their regression coefficient values are summarized in Table 2:

Predictor	Estimate	p-value	Significance [1]
(Intercept)	-0.0026054	0.2802696	
share_tertiary	0.0231545	0.0000000	***
share_secondary	-0.0111015	0.0104648	*
share_20_29	-0.0346547	0.0004859	***
share_30_39	0.0218416	0.0005671	***
share_40_49	0.0428472	0.0007255	***

[1] Using convention: 0 = ***, 0.001 = **, 0.01= *, 0.05 = .

Table 2 Regression Coefficients

The predictors selected using the stepwise regression approach can be described as:

- *share_tertiary* – ratio of population with tertiary education to population over 15 years of age
- *share_secondary* – ratio of population with secondary education (only) to population over 15 years of age
- *share_20_29* – ratio of population in age bracket 20 to 29 years of age to total population
- *share_30_39* – ratio of population in age bracket 30 to 39 years of age to total population
- *share_40_49* – ratio of population in age bracket 40 to 49 years of age to total population

The coefficients identified via stepwise selection in Table 2 can be interpreted as belonging to two groups: the first related to education – with the share of population with tertiary education leading to higher EV penetration, and share of population with secondary education only leading to lower EV penetration (yet with lower magnitude of the effect).

The second group relates to age structure: share of population in early productive age (20–29 years) drives EV adoption down, and share of population in middle to upper middle productive age (30–39 and 40–49) drives EV adoption up. Other socioeconomic factors, including those related to employment status, nationality and religion, did not have a significant effect on EV adoption as measured by the BIC criterion.

A possible common theme behind the two groups of socioeconomic predictors identified is income. While average income was not one of the census variables considered it is known to correlate with both education level and age [7] – with average wages increasing with education, and peaking in the 40 to 44 age bracket.

The R^2 statistic for our proposed model is 0.3485121, meaning our model does not fully explain the variance observed. On the other hand the F statistic is 21.39792, indicating a highly significant relationship.

Based on visual overview shown in Figure 1 we formed a hypothesis of spatial heterogeneity of retail EV penetration – given the noticeable clusters of high penetration in Prague, Brno and their outskirts. To formally validate our hypothesis a test was performed using Lagrange multiplier diagnostics for spatial dependence. The test statistic of 288.4171 at 1 degree of freedom strongly indicates a spatial dependence.

In addition the observed value of Moran's I statistic for EV penetration was 0.2615302, with z-score 6.243977 indicating a significant autocorrelation at p-value $2.1329152 \times 10^{-10}$.

Compared to the Moran I test under randomisation for original values, which showed strong autocorrelation, the same test for residuals after linear modelling for the 5 socioeconomic predictors displays much less spatial dependence, with Moran's I statistic of 0.01352744, with z-score 0.4261943. This implies p-value 0.3349831, leading us to reject a hypothesis of spatial autocorrelation of model residuals.

Finally we performed the Lagrange multiplier diagnostics for spatial dependence test on the model fitted for 5 socioeconomic predictors. Test statistic of 0.09503605 at 1 degree of freedom implies p-value 0.7578699, again leading us to reject a hypothesis of spatial dependence.

Thus we confirmed that the socioeconomic model describes the observed data efficiently and that it has removed the observed spatial dependency from EV penetration data, leaving only random effect.

4 Conclusion

We propose a model explaining the dynamics of Electric Vehicles in the Czech Republic retail segment based on socioeconomic predictors at the level of 206 Czech administrative units. The model identified 5 significant predictors (two related to education, and three related to age structure). The identified predictors are variables

known to correlate with personal income. While our model did not explain the observed differences in EV adoption over Czech ORPs fully, it did remove the effect of spatial autocorrelation from the data, leaving only random effect.

5 Acknowledgement

This contribution was supported by the Prague University of Economics and Business project IGS F4/24/2023.

6 References

[1] 341/2014 Sb. Vyhláška o schvalování technické způsobilosti a o technických podmínkách provozu vozidel na pozemnních komunikacích.

[2] Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity, *Geographical Analysis*, vol. 20, no. 1, 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x.

[3] Bivand, R. & Pebesma, E. (2023). *Spatial Data Science with applications in R*, Chapman & Hall/CRC. https://doi.org/10.1201/9780429459016

[4] Bouchal, P. (2022). *czso: Use Open Data from the Czech Statistical Office in R.* Available at: https://CRAN.R-project.org/package=czso [cited 2023-08-15]

[5] Bryła, P., Chatterjee, S. & Ciabiada-Bryła, B. (2022). Consumer Adoption of Electric Vehicles: A Systematic Literature Review, *Energies*, vol. 16, no. 1. https://doi.org/10.3390/en16010205.

[6] Český statistický úřad (2011). *Výsledky sčítání lidu, domů a bytů 2011*. Available at: https://www.czso.cz/csu/czso/otevrena_data_pro_vysledky_scitani_lidu_domu_a_bytu_2011_sldb_2011 [cited 2023-08-15]

[7] Český statistický úřad (2022). *Struktura mezd zaměstnanců – 2021*. Available at: https://www.czso.cz/csu/czso/struktura-mezd-zamestnancu-2021 [cited 2023-08-15]

[8] Hardman, S. et al. (2018). A review of consumer preferences of and interactions with electric vehicle charging infrastructure *Transportation Research Part D: Transport and Environment*, vol. 62, 508–523. https://doi.org/10.1016/j.trd.2018.04.002.

[9] Jenn, A., Springel, K. & Gopal, A. R. (2018). Effectiveness of electric vehicle incentives in the United States, *Energy Policy*, vol. 119, 349–356. https://doi.org/10.1016/j.enpol.2018.04.065.

[10] Kester, J., Noel, L., Zarazua de Rubens, G. & Sovacool, B. K. (2018). Policy mechanisms to accelerate electric vehicle adoption: A qualitative review from the Nordic region, *Renewable and Sustainable Energy Reviews*, vol. 94, 719–731. https://doi.org/10.1016/j.rser.2018.05.067.

[11] Lacko, J. (2023). RCzechia: Spatial Objects of the Czech Republic, *Journal of Open Source Software*, vol. 8, no. 83, 5082. https://doi.org/10.21105/joss.05082.

[12] Lumley, T. based on Fortran code by Miller, A. (2020). *leaps: Regression Subset Selection.* Available at: https://CRAN.R-project.org/package=leaps [cited 2023-08-15]

[13] Ministerstvo dopravy ČR (2023). *Registr silničních vozidel* Available at: https://www.mdcr.cz/Statistiky/Silnicni-doprava/Centralni-registr-vozidel [cited 2023-08-15]

[14] Parry, J. (2023). *sfdep: Spatial Dependence for Simple Features.* Available at: https://CRAN.R-project.org/package=sfdep [cited 2023-08-15]

[15] R Core Team (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Available at: https://www.R-project.org/ [cited 2023-08-15]

[16] Schulz, F. & Rode, J. (2022). Public charging infrastructure and electric vehicles in Norway, *Energy Policy*, vol. 160. https://doi.org/10.1016/j.enpol.2021.112660.

[17] Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, vol. 6, no. 2, 461–464. https://doi.org/10.1214/aos/1176344136.

[18] Slabá, R. & Houšť, R. (2022). *Ročenka dopravy České Republiky 2021,* Ministerstvo dopravy ČR, Available at: https://www.sydos.cz/cs/rocenka-2021/index.html [cited 2023-08-15]

[19] Ščasný, M., Zvěřinová, I. & Czajkowski, M. (2018). Electric, plug-in hybrid, hybrid, or conventional? Polish consumers' preferences for electric vehicles, *Energy Efficiency*, vol. 11, no. 8, 2181–2201. https://doi.org/10.1007/s12053-018-9754-1.

Analysis of Commercial Property Prices on the Czech Market

Michaela Matoušková[1]

Abstract. The development of the commercial real estate market plays an important role in the financial stability of the country, especially through the level of credit of companies operating in this area. This article analyses the price development of the Commercial Real Estate Capital Value Index (CRECVI) in the Czech market along with selected economic macro aggregates. Specifically, the market for production and storage facilities was analysed.
Within the framework of the cointegration analysis, an ADL model is constructed to describe how prices of the production and storage facilities react to changes in selected macroeconomic indicators. From the ADL model, an error correction model is then obtained by recalculation, which separately describes the short-run and long-run relationships between the time series. The results of the analysis show that the prices of warehouse properties are mainly influenced by the evolution of gross domestic product, unemployment and the level of inflation and interest rates.

Keywords: cointegration analysis, CRECVI index, ADL model, commercial real estate, EC model

JEL Classification: C51, F63
AMS Classification: 62P20, 91B02

1 Introduction

Investing in real estate has become a society-wide phenomenon in recent years and is becoming more and more widespread. Commercial real estate is often part of investment fund portfolios. Specifically, they are included in qualified investor funds or in a special real estate fund. These funds are generally classified as more risky, mainly due to their reduced liquidity and less frequent valuation of units and underlying assets. [14]

A large amount of space is devoted to the development of residential property prices in the literature. The analysis of commercial property prices is more in the background and focuses mainly on the office market. [8] Therefore, this paper focuses on the price development of production and storage facilities on a dataset that is not publicly available. The aim is thus to provide new information in the Czech Republic. Describing and modelling the evolution of production and storage facilities prices can provide insight into the history and dynamics of this market. Based on the econometric model obtained, it is then also possible to predict subsequent price developments.

Commercial real estate market is a very specific market as prices are always linked to a concrete area and also because of heterogeneous range of properties. A typical feature is the inelasticity of supply versus demand. The monitored real estate transactions are focused only on the prime sector of the market, representing only the most lucrative properties. The rest of the market, which does not meet the prime property standard, is thus not included in the analysis. However, the advantage of tracking only prime real estate is comparability with other countries. Thus, these analyses mainly cover capitals and other large cities. [8] [13]

1.1 Development of the Industrial Real Estate Market

According to Savills' analysis [10], there has been a significant increase in the total stock of industrial space for lease during the first three quarters of 2022. With the same period in 2021, this represented a 161% increase. Although the area of completed industrial space reached very high numbers, the vacancy rate in Q3 was the lowest in recent years. In Prague, the vacancy rate was below 1% at the end of 2022 and the average for the whole country was 1.2%.

The volume of properties under construction rose to 1.33m sqm from 1.12m sqm at the beginning of the year, representing 12% of supply. More than half of this pipeline was already pre-leased at the end of the year.

[1] Technical university of Liberec, Department of Finance and Accounting, Studentská 1402/2, Liberec, 461 17, Czech Republic, michaela.matouskova@tul.cz.

The Karlovy Vary Region had the largest share of the pipeline (24%) with 319,700 sqm. This was due to the construction of the largest industrial hall in the Czech Republic, which has an area of approximately 233,700 m2. Other regions with the largest share of warehouse and production space construction at the end of 2022 included the Ústí nad Labem Region (14%), the South Moravian Region (11%) and the Pilsen Region (11%).

The total estimated volume of completed industrial lease space for 2022 was up to 1.33 million sqm, the highest area of completed facilities to date. This is a year-on-year increase of 96%. [11]

2 Methodology

For the purposes of this analysis, commercial real estate is defined as real estate that is used solely for business purposes. The commercial real estate market consists mainly of the following segments: office buildings, industrial and warehouse buildings, retail premises and hotels. The cointegration analysis in this paper will focus specifically on the industrial and warehouse market.

2.1 Cointegration

As stated by Arlt [1] cointegration is one way of classifying economic time series where time series are divided into short memory and long memory series. Time series cointegration was first studied in the early 1980s by C. W. J. Granger. This method is based on the problem of integrated processes, which had already been addressed by G. Box and G. Jenkins.

According to Arlt [2], when modelling multivariate economic time series, it is useful to distinguish between short-run and long-run relationships, as short-run relationships between time series exist only in a relatively short period and fade over time. These short-term relationships occur in non-stationary time series, which are characterized by their short memory. The second type of relationship is long term in nature and persists over time.

Long-term relationships between time series are closely related to the concept of equilibrium, which can be understood as a steady state. The system is continuously attracted to this equilibrium state. However, the system is subject to shocks and is never directly in equilibrium, but may be in a long-term equilibrium towards which it converges over time. A time series is in long-term equilibrium if it does not diverge in the long run. Therefore, an analysis of the long-run relationship between time series can only be performed for non-stationary time series that share a common stochastic trend. These time series are then considered to be cointegrated. [2]

According to Brooks [6], if the time series have a different direction of trend, the analysis of the relationship produces a condition referred to as apparent regression. An apparent regression is considered to be a situation where there are time series that are unrelated. However, it is possible to obtain statistically significant estimates of the parameters of the regression function using the least squares method. Thus, the index of determination, t-tests and F-test will indicate the appropriateness of using the model. Thus, the time series cointegration test also serves as an indicator of true and apparent regression.

Testing for cointegration in univariate models can be based on testing the stationarity of the residuals. The residuals required for testing are estimated using the least squares method, where one series is considered as the explanatory variable and the other series as the explanatory variables. This is based on a regression model of the form:

$$Y_t = \beta X_t + a_t \tag{1}$$

For this testing, the Augmented Dickey-Fuller test (ADF test) is mainly used. This tests the hypothesis that the time series are not cointegrated, i.e. that the non-systematic component is of type I(1). In this case, it is an apparent regression. If the ADF test shows that the residuals are stationary, it is a true cointegrating regression.

If the unsystematic component of the model represents white noise, a simple linear regression is sufficient to capture the relationships. If the autocorrelation of the non-systematic component is evident, a lagged explanatory variable model, the Autoregressive Distributed Lag model (ADL model), is used.

In the diagnostic check of the model, according to Arlt [3], it is necessary to test whether the nonsystematic component exhibits normality, homoskedasticity and is not autocorrelated. To assess the normality of the unsystematic component, the Jarque-Bera test is used, which is based on simultaneously testing the skewness and skewness of the unsystematic component. To test the homoskedasticity of the unsystematic component of the model, the ARCH ("AutoRegressive Conditional Heteroskedasticity") effect is tested. This test consists of creating an artificial regression where the explanatory variable is the square of the residuals and the explanatory variable is the square of the residuals in lag q.

According to Pagan [10], the Breusch-Godfrey LM test can be used to assess the autocorrelation of the unsystematic component. This test verifies the serial interdependence of the random components in the model, where the null hypothesis states that there is no autocorrelation in the model. The test consists of creating an artificial variable where the explanatory variable is at, the explanatory variables are *at-1*, ..., *at-p* and the explanatory variables of the model are $y = \beta X + a_t$.

After identifying cointegrated relationship between variables, we can employ Autoregressive Distributed Lag model ADL [10]. Model has form ADL(1,1)

$$Y_t = c + a_1 Y_{t-1} + \beta 1 X_t + \beta 2 X_{t-1} + a_t, \tag{2}$$

According to Brooks [6], a group of cointegrated time series can then be described by an error-correction ("error-correction", EC) model that can distinguish between short-run and long-run relationships. The model can be vritten in the form:

$$\Delta Y_t = c + \beta 1 \Delta X_t + \gamma (Y_{t-1} - \beta X_{t-1}) + a_t, \tag{3}$$

where $\beta = (\beta_1 + \beta_2)/(1 - \alpha_t)$ and $\gamma = \alpha_1 - 1$. The long-run relationship is expressed by a regressor $(Y_{t-1} - \beta X_{t-1})$ that contains a long-run multiplier β. The given regressor forms a component of the EC model. The remaining part of the model represents the short-term relationship between the time series. The parameter γ denotes the degree to which the short-run relationship differs from the long-run relationship. It can also be interpreted as the strength with which the equilibrium relationship between time series. If the parameter is equal to zero, there is no cointegration between the time series. [12]

As stated by Arlt [4], when modelling the relationship between I(1) type time series, it is not appropriate to stationarize them by differencing for cointegrated time series. When differencing these series, important information contained in the EC model is lost. The importance of the EC model lies in the modelling of time series transformed by differencing and the original untransformed time series. Thus, in the EC model, it is possible to simultaneously capture the short-term relationships between processes, which are the relationships between differentiated processes, and the long-term relationships between undifferentiated processes.

3 Data

The CRECVI Index was provided for the purposes of this analysis by the consultancy Cushman & Wakefield. Quarterly data for the explanatory variables were obtained from the Czech Statistical Office.

3.1 CRECVI Index

The C&W CRECVI (Commercial Real Estate Capital Value Index) is compiled by the consultancy Cushman & Wakefield. The index is limited to the Czech Republic and shows year-on-year changes in the capital value of a portfolio of commercial real estate (prime Prague offices, shopping centers and logistics parks). The index is based on quarterly "prime rents" and "prime yields" for selected commercial real estate markets. In this paper, the index focuses only on industry - production and storage facilities (CRECVI_IND, Q1/2015 = 100).

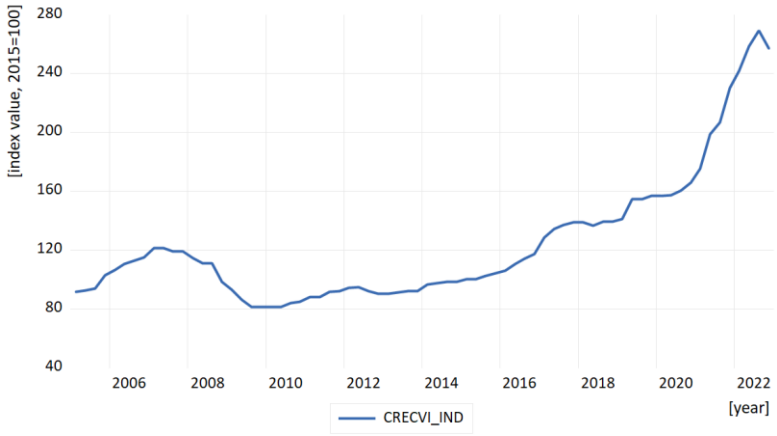


Figure 1 Development of the CRECVI – logistics parks

The cointegration analysis is based on quarterly data for the period I/2005 - IV/2022. In addition to the CRECVI index (production and storage facilities), it also includes the macroeconomic variables, listed below:

- GDP (GDP);
- Consumer Price Index (CPI);
- Interest rate for commercial real estate (INTRST_RATE);
- Unemployment (UNEMPL).

These variables were used to model the short- and long-term effects on the commercial real estate capital value index, which is the explanatory variable.

4 Empirical Results

The time series modelling was done in EViews 13. Except for the explanatory variable of the CRECVI index, all the time series exhibited seasonality. The C-X13 ARIMA method was chosen to identify the seasonal component and subsequently adjust the time series. After adjustment, the time series are denoted by the ending *SA* in the analysis.

Based on the results of the augmented Dickey-Fuller unit root test, the time series are non-stationary, type I(1). The model residuals obtained from the linear regression are stationary. The unit root test confirmed that the time series are cointegrated.

The statistically significant ADL model is presented in Table 1.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CRECVI_IND(-1)	1,067683	0,028438	37,54451	0,0000
CPI_SA(-1)	-0,618222	0,14449	-4,27866	0,0001
GDP_SA	4,30E-05	1,16E-05	3,695263	0,0005
INTRST_RATE_SA	3,476577	0,830728	4,184975	0,0001
INTRST_RATE_SA(-1)	-5,521994	0,824733	-6,69549	0,0000
UNEMPL_SA(-1)	2,286855	0,425002	5,380805	0,0000

Table 1 ADL model

After fitting the ADL model, a diagnostic tests was performed and the results are presented in Table 2.

Model diagnostics	Stat.	Prob.
R^2	0,992783	-
Durbin-Watson Stat.	2,3850	-
Breusch-Godfrey	1,740111	0,1838
Jarque-Bera	17,55634	0,0001
ARCH	3,360079	0,0712

Table 2 Diagnostic tests

A correlogram was used to show the progression of ACF and PACF, which did not show autocorrelation. Also, the DW statistic did not show autocorrelation in the model. The ARCH test shows, that the non-systematic component of the model is homoscedastic. Further, in validation of the model fit, Breusch-Godfrey test was conducted which shows that the non-systematic component of the model is not autocorrelated. Jarque-Bera test was used to test the normality of the model, according to which the non-systematic component of the model is not normally distributed. As the histogram in Figure 2 shows, the test failed due to slight skewness of the data. The least squares method is very sensitive to the failure of normality. In the context of this analysis, the slight skewness of the data can be considered within tolerance.

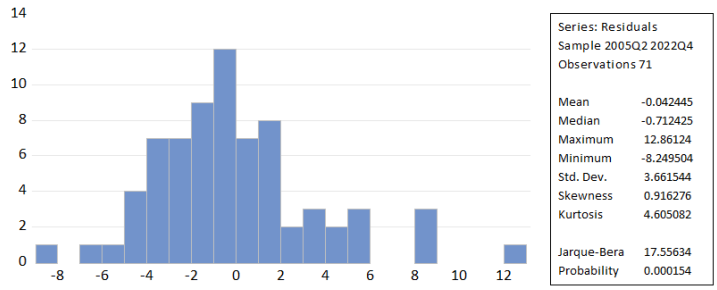


Figure 2 Jarque-Bera test

The resulting ADL model passed the diagnostic check and can be written as:

$$\widehat{CRECVI\_IND}_t = 1{,}0677\ CRECVI\_IND_{t-1} - 0{,}6182\ CPI\_SA_t + 4{,}30\text{E}-05\ GDP\_SA_t + 3{,}4766\ INTRST\_RATE\_SA_t - 5{,}522\ INTRST\_RATE\_SA_{t-1} + 2{,}2869\ UNEMPL\ SA_{t-1} \quad (3)$$

According to the resulting model, the commercial real estate index at time *t* depends on its value at time *t-1*. The equation shows that at time *t* the prices of industrial and warehouse properties are influenced by consumer price index, interest rate and gross domestic product. The interest rate and unemployment variables then depend on the price at time *t-1*.

According to the model, gross domestic product has a direct proportional effect on commercial property prices. GDP growth is reflected in the growth of demand for real estate, which puts upward pressure on prices. At lag t-1, real estate prices are inversely affected proportionally by interest rates on commercial real estate mortgage loans. As interest rates fall, properties become more affordable and prices increase. The consumer price index came out statistically significant, indicating that as inflation falls, mortgage interest rates fall. Low interest rates are pushing up demand for property, which is driving up house prices in the face of a shortage of supply.

From the ADL model we obtained the EC error correction model after recalculation:

$$\Delta\widehat{CRECVI\_IND}_t = -0{,}6182\ CPI\_SA_t + 4{,}30\text{E}-05\ GDP\_SA_t + 3{,}4766\ INTRST\_RATE\_SA_t + 0{,}0677(CRECVI\_IND_{t-1} + 81{,}56573\ INTRST\_RATE\_SA_{t-1} - 33{,}7799\ UNEMPL\ SA_{t-1}) \quad (4)$$

From the EC model we can see, how fast the system tends to the long-term equilibrium as measured by parameter γ (0,0677) and the long-term relationship measured by multiplier β1 (81,56573) and β2 (-33,7799). The long-term relations is expressed by the term EC in equation:

$$EC_t = CRECVI\_IND_{t-1} + 81{,}56573\ INTRST\_RATE\_SA_{t-1} - 33{,}7799\ UNEMPL\ SA_{t-1} \quad (5)$$

5 Conclusion

The article analyses the cointegration relationship between index of commercial properties, specifically production and storage facilities, in the Czech market and selected macroeconomic aggregates. The ADL model was constructed to describe how production and storage facilities prices react to changes in economic. From the ADL model, an error correction model is then obtained by recalculation, which separately describes the short-run and long-run relationships between the time series. The results of the analysis show that the prices of production and storage facilities are mainly influenced by the evolution of gross domestic product, unemployment and the level of inflation and interest rates.

Further research in this area could be focused on the analysis during the pandemic of COVID. A further extension of this research can also be to compare the evolution of the commercial real estate index across EU countries.

Acknowledgements

This research was created in accordance with the institutional support for the conceptual development of the Faculty of Economics of the University of Liberec Project: Internal grant competition called “Dopady pandemie COVID-19 na finanční výkonnost vybraných podnikatelských subjektů” and Project Student grant competition “Ekonometrické modelování indexu kapitálové hodnoty komerčních nemovitostí na českém trhu”.

References

[1] Arlt, J. & Arltová M. (2007). *Ekonomické časové řady*. Grada Publishing.
[2] Arlt, J. & Arltová M. (2003). *Finanční časové řady*. Grada Publishing.
[3] Arlt, J., Arltová M. & Rublíková, E. (2002). *Analýza ekonomických časových řad s příklady*. 2. vyd. Skripta VŠE Praha.
[4] Arlt, J. (1999). *Moderní metody modelování ekonomických časových řad*. Grada Publishing.
[5] Arlt, J. (1997). Kointegrace v jednorovnicových modelech. *Politická ekonomie, 45(5)*, 733-746.
[6] Brooks, Ch. (2008). *Introductory Econometrics for Finance*. Cambridge University Press.
[7] Hlaváček, M. & Komárek L. (2010). Rovnovážnost cen nemovitostí v České republice. *Politická ekonomie, 3*, 326-342.
[8] Hlaváček, M., Novotný O. & Rusnák M. (2016). Analýza cen komerčních nemovitostí v zemích střední Evropy. *Politická ekonomie, 1*, 3-18.
[9] Hušek, R. (2008). *Ekonometrická analýza*. Oeconomica VŠE.
[10] Pagan, A. (2009). Time series behavior and dynamic specification. *Oxford Bulletin of Economics and Statistics. 47(3)*, 199-211.
[11] Savills (2022). *European Logistics Outlook*. Savills Commercial Research.
[12] Stock, J. H. & Watson, M.W. (2001). Vector Autoregressions. *Journal of Economic Perspectives, 15(4)*, 101-115.
[13] Stoklasová, R. (2018). Default rate in the Czech Republic depending on selected macroeconomic indicators. *E&M Economics and Management, 21(2)*, 69-82.
[14] Wheaton, W. (1999). Real estate „cycles“: some fundamentals. *Real estate economics, 27(2)*, 209-230.

Distributionally Robust Fixed Interval Scheduling with Heterogeneous Machines under Uncertain Finishing Times

Monika Matoušková[1]

Abstract. We deal with operational fixed interval scheduling problems where start times are given and the actual finishing times can be influenced by random delays. We further consider heterogeneous case, i.e., multiple job and machine types. And we assume that the multivariate distribution of delays follows an Archimedean copula. We consider the highest worst-case probability that the schedule remains feasible, where given proportion of marginal distributions of delays are stressed. This problem has a reformulation containing a commonly used risk measure. We implement a decomposition algorithm and compare it with MIP solver.

Keywords: Stochastic optimization, Fixed interval scheduling, Heterogeneous machines, Risk measures

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

We are dealing with distributionally robust fixed interval scheduling (FIS) problem with heterogeneous machines and random delays. We have multiple machine and job types. The times when each job begins and ends are given. But there can be random delays that depend on machine type where particular job is assigned to. The task is to assign each job to some machine and to process at most one job at a time on each machine. Further, we are looking for the robust FIS problem solution, i.e., a schedule with highest worst-case probability that the schedule remains feasible. We show a possible problem reformulation and present two ways of problem solution. First is using MIP solver and then using a decomposition algorithm based on golden-section search algorithm from [2]. We are extending the distributionally robust FIS problem for homogeneous machines which was solved in [1].

One application of the problem that we could consider is the gate assignment problem. In this practical problem, the incoming flights have to be assigned to available gates of an airport. Different aircrafts might need to be assigned to different gates, which is why the heterogeneous case is useful. Moreover, there can be some proportion of flights where worse delays occur. This situation can be included in the FIS problem by setting the ambiguity set of the delay distributions appropriately. This problem is than a FIS problem where we maximize the worst-case probability that the flight assignment to the gates remains feasible.

2 Problem Formulation

We have set of jobs $\mathcal{J} = \{1, \ldots, J\}$ and set of machines $C = \{1, \ldots, C\}$. For machine $c \in C$ its class is denoted by b_c and set of all job classes is $\mathcal{B} = \{b_c : c \in C\} = \{1, \ldots, B\}$, where B is a number of machine classes.

For job $j \in \mathcal{J}$ its type is denoted by a_j. Let A be a number of all job types. Symbol $\mathcal{J}_b$ denotes a set of jobs, which can be assigned to a machine in class $b \in \mathcal{B}$, their number is J_b. One job of a given type can be processed by one or more machine classes. Any machine from class $b \in \mathcal{B}_j$ can process job $j \in \mathcal{J}$. A number of machines from class $b \in \mathcal{B}$ is denoted by C_b.

For each job $j \in \mathcal{J}$ we consider its start time s_j, a set of all job beginnings is denoted by $\mathcal{T} = \{s_1, \ldots, s_J\}$. It is assumed that the random ending of job $j \in \mathcal{J}$ can be written as $f_j(\xi) = f_j + D_{jb}(\xi)$ depending on the machine type it is assigned to, where f_j is a prescribed end time and $D_{jb}(\xi)$ is a random delay. The random delay is a random variable with a known distribution that is induced by a probability space $(\Xi, \mathcal{F}, \mathsf{P})$.

The robust FIS problem with heterogeneous machines can be written as a problem of maximization of the worst-case

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

probability that the schedule remains feasible:

$$\begin{aligned}\max_{x}\min_{\mathsf{P}\in\mathcal{P}}\ \mathsf{P}\left(\xi\in\Xi:\ \sum_{j\in\mathcal{J}_b:\ s_j\le t<f_j(\xi)} x_{jb}\le C_b,\quad b\in\mathcal{B},\quad t\in\mathcal{T}\right)\\ \text{s.t.}\ \sum_{j\in\mathcal{J}_b:\ s_j\le t<f_j} x_{jb}\le C_b,\quad b\in\mathcal{B},\quad t\in\mathcal{T},\\ \sum_{b\in\mathcal{B}_j} x_{jb}=1,\quad j\in\mathcal{J},\\ x_{jb}\in\{0,1\},\quad j\in\mathcal{J}_b,\quad b\in\mathcal{B},\end{aligned}\tag{1}$$

where $\mathcal{P}$ is an ambiguity set. The binary variables x_{jb} indicate whether job $j\in\mathcal{J}$ is assigned to a machine of type $b\in\mathcal{B}_j$. In the objective function there is a probability that each machine type $b\in\mathcal{B}$ processes maximum of C_b jobs at a time, i.e., it never processes more jobs than there are available machines of this type, if we take into account the random delays. And it is sufficient to look only at start time of each job. The first set of constraints gives what is inside of the probability but only for prescribed ending times of jobs. So no machine processes more than one job at a time considering the prescribed end times. The second set of constraint gives that each job is processed by exactly one machine type, and so by exactly one machine.

2.1 Assumptions

Further, we will assume that

$$\mathsf{P}(D_{jb}(\xi)\le x_{jb}, j\in\mathcal{J}_b, b\in\mathcal{B})=\psi^{-1}\left(\sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{J}_b}\psi(F_{jb}(x_{jb}))\right),$$

i.e., the multivariate distribution of random delays follows an Archimedean copula with generator ψ, where the function $\psi:[0,1]\to[0,\infty]$ has to be continuous decreasing and has to satisfy $\psi(1)=0,\ \lim_{x\to 0_+}\psi(x)=\infty$.

Now, consider that the probability that there will be no delay of job j assigned to machine of type b is $p_{jb}\in[0,1]$. We use exponential distribution with parameter λ_{jb} for length of the delay of job j assigned to machine of type b. Marginal cumulative distribution function for the pair of indices j, b is then $F_{jb}(x)=p_{jb}+(1-p_{jb})(1-\mathrm{e}^{-\lambda_{jb}x})$, $x\ge 0$. This assumption is not restrictive, the following holds also for different choice of the distribution function, we use it further in the numerical simulation.

For the ambiguity set $\mathcal{P}$ we suppose that we have some cdf $\tilde{F}_{jb}(x)<F_{jb}(x)$, $x\ge 0$, $\forall j\in\mathcal{J}_b$, $b\in\mathcal{B}$, which is used for stressing. In our case, the stressing distribution function $\tilde{F}_{jb}$ can be taken such that p_{jb} is lowered or λ_{jb} is enlarged (or both). We will stress delays of Γ jobs, so all F_{jb} will be stressed for each $b\in\mathcal{B}_j$ for a stressed job $j\in\mathcal{J}$. And the set equals

$$\mathcal{P}=\left\{\begin{array}{c}\text{delays of }\Gamma\text{ jobs from }\{1,\dots,J\}\text{ are stressed, i.e., pick }\tilde{j}_1,\dots,\tilde{j}_\Gamma\text{ jobs whose delays will be stressed,}\\ \text{marginal distributions }F_{jb}\text{ are stressed to }\tilde{F}_{jb}\,\forall j\in\{\tilde{j}_1,\dots,\tilde{j}_\Gamma\}\,\forall b\in\mathcal{B}_j,\\ \text{remaining marginal distributions }F_{jb}\text{ are unchanged,}\\ \text{joint distribution follows an Archimedean copula with generator }\psi\end{array}\right\}$$

with $0<\Gamma<J-C$, because stressing last job on a machines means no change, because the penalty is zero for such a job as it can cause no schedule infeasibility.

2.2 Problem Reformulation

Under the stated assumptions, the robust operational FIS problem has reformulation based on network flow formulation, now with Conditional Value at Risk in our objective function. The Conditional Value at Risk (CVaR) is a commonly used risk measure which can be written using the following minimization formula according to [3]

$$\mathrm{CVaR}_\alpha(Z)=\min_\theta\left\{\theta+\frac{1}{(1-\alpha)S}\sum_{s=1}^{S}\max\{Z^s-\theta,0\}\right\},$$

where Z is a random loss variable with S equiprobable realizations Z^S.

But before we get to the reformulation, we have to introduce further notation. We define a set of edges for machine class $b \in \mathcal{B}$ denoted by $\bar{E}_b$, which contains all pairs of jobs that can be assigned to the same machine from class $b \in \mathcal{B}$ consecutively, i.e., $\{j,k\} \in \bar{E}_b$, if $f_j \leq s_k$ and $j,k \in \mathcal{J}_b$. We denote by E_b a set of edges containing all edges from $\bar{E}_b$ but also edges $\{0, J+1\}$ and $\{0, j\}, \{j, J+1\}$ for $j \in \mathcal{J}_b$.

For $\{j,k\} \in \bar{E}_b$, $b \in \mathcal{B}$, we consider penalizations $q_{jkb} = \psi(\mathsf{P}(D_{jb}(\xi) \leq s_k - f_j)) = \psi(F_{jb}(s_k - f_j))$, for stressed case $\tilde{q}_{jkb} = \psi(\mathsf{P}(\tilde{D}_{jb}(\xi) \leq s_k - f_j)) = \psi(\tilde{F}_{jb}(s_k - f_j))$, where ψ is a generator function of an Archimedean copula. Their difference is denoted by $\Delta_{jkb} = \tilde{q}_{jkb} - q_{jkb} > 0$. And penalizations for $\{j,k\} \in E_b \setminus \bar{E}_b$ are $q_{jkb} = \tilde{q}_{jkb} = \Delta_{jkb} = 0$.

Next, we define $E_{\cdot jb} = \{i \in \mathcal{J}_b \cup \{0\} : \{i,j\} \in E_b\}$ a set of possible predecessors of job j on machine class b, $E_{j\cdot b} = \{k \in \mathcal{J}_b \cup \{J+1\} : \{j,k\} \in E_b\}$ a set of possible successors of job j on machine class b, $j \in \mathcal{J}_b$, $b \in \mathcal{B}$. If we have $b \in \mathcal{B}$ and $j \in \mathcal{J} \setminus \mathcal{J}_b$, then $E_{\cdot jb} = E_{j\cdot b} = \emptyset$.

Theorem 1. *Let the assumptions in subsection 2.1 hold. Then the robust FIS problem (1) has the following reformulation*

$$\min_y \sum_{b\in\mathcal{B}} \sum_{\{j,k\}\in\bar{E}_b} q_{jkb} y_{jkb} + \Gamma \mathrm{CVaR}_\alpha(Z(y))$$

$$\text{s.t.} \sum_{j\in\mathcal{J}_b\cup\{J+1\}} y_{0jb} = C_b, \quad b \in \mathcal{B}, \tag{2a}$$

$$\sum_{i\in E_{\cdot jb}} y_{ijb} = \sum_{k\in E_{j\cdot b}} y_{jkb}, \quad j \in \mathcal{J}_b, \quad b \in \mathcal{B}, \tag{2b}$$

$$\sum_{j\in\mathcal{J}_b\cup\{0\}} y_{j(J+1)b} = C_b, \quad b \in \mathcal{B}, \tag{2c}$$

$$\sum_{b\in\mathcal{B}_j} \sum_{k\in E_{j\cdot b}} y_{jkb} = 1, \quad j \in \mathcal{J}, \tag{2d}$$

$$y_{jkb} \in \{0,1\}, \quad \{j,k\} \in E_b, \quad b \in \mathcal{B}, \tag{2e}$$

where $\alpha = 1 - \Gamma / \sum_{b\in\mathcal{B}} |\bar{E}_b|$. *And* $Z(y)$ *is loss random variable with equiprobable realizations* $\Delta_{jkb} y_{jkb}$, $\{j,k\} \in \bar{E}_b$, $b \in \mathcal{B}$. *The binary variable* y_{jkb} *indicates whether job* $k \in \mathcal{J}$ *is planned after job* $j \in \mathcal{J}$ *on the same machine of type* $b \in \mathcal{B}$.

Proof. Under the assumption of Archimedean copula distribution of the random delays with generator function ψ, the probability of the schedule remaining feasible in the worst-case can be written as

$$\mathsf{P}\left(\begin{array}{c} \tilde{D}_{jb}(\xi) \leq s_k - f_j \text{ if job } j \text{ has successor } k \text{ on a machine of type } b \text{ and delay of job } j \text{ is stressed,} \\ D_{jb}(\xi) \leq s_k - f_j \text{ if job } j \text{ has successor } k \text{ on a machine of type } b \text{ and delay of job } j \text{ not stressed,} \\ \tilde{D}_{jb}(\xi), D_{jb}(\xi) \leq \infty \text{ otherwise, } \{j,k\} \in \bar{E}_b,\ b \in \mathcal{B} \end{array}\right)$$

$$= \psi^{-1}\left(\sum_{b\in\mathcal{B}} \sum_{\{j,k\}\in\bar{E}_b} \psi((1-\tilde{z}_{jkb}) p_{jkb} + \tilde{z}_{jkb} \tilde{p}_{jkb})\right) = \psi^{-1}\left(\sum_{b\in\mathcal{B}} \sum_{\{j,k\}\in\bar{E}_b} ((1-\tilde{z}_{jkb}) q_{jkb} + \tilde{z}_{jkb} \tilde{q}_{jkb}) y_{jkb}\right) \tag{3}$$

using the fact that $\tilde{z}$'s are binary in the last equality. Where we define the probabilities

$$p_{jkb} = \begin{cases} \mathsf{P}(D_{jb}(\xi) \leq s_k - f_j), & \text{if job } j \text{ has successor } k \text{ on a machine of type } b, \\ \mathsf{P}(D_{jb}(\xi) \leq \infty) = 1, & \text{otherwise,} \end{cases}$$

and similarly $\tilde{p}_{jkb}$ with $\tilde{D}_{jb}$ and $\mathsf{P}(\tilde{D}_{jb}(\xi) \leq s_k - f_j) = \tilde{F}_{jb}(s_k - f_j)$, then we define the binary variables

$$y_{jkb} = \begin{cases} 1, & \text{if job } j \text{ has successor } k \text{ on a machine of type } b, \\ 0, & \text{otherwise,} \end{cases}$$

and also

$$\tilde{z}_{jkb} = \begin{cases} 1, & \text{if delay of job } j \text{ is stressed,} \\ 0, & \text{otherwise,} \end{cases}$$

for $\{j,k\} \in \bar{E}_b$, $b \in \mathcal{B}$.

The problem (1) has the reformulation given by minimizing the probability in (3) subject to the constraints (2a) - (2e) and $\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}\tilde{z}_{jkb}=\Gamma,\ \tilde{z}_{jkb}\in\{0,1\},\ \{j,k\}\in\bar{E}_b,\ b\in\mathcal{B}$.

Let us now comment the constraints of the problem. Constraints (2a) - (2c) are modelling the flow. Particularly, the set of constraints (2a) and (2c) ensure that we do not use more machines of certain class than we have. Set of constraints (2b) ensure that considering the prescribed ending times, no jobs overlap. And (2a) - (2c) also secure that we use the corresponding machine class for each job. The constraint (2d) gives that each job is processed by exactly one machine type, thus by exactly one machine as those variables are binary. Due to this constraint the constraint matrix is not totally unimodular. The constraint on $\tilde{z}$ gives that Γ jobs are stressed. Consider we have y feasible, thus $\sum_{b\in\mathcal{B}_j}\sum_{k\in\bar{E}_{j,b}} y_{jkb}\leq 1$ for each j, so at most one coefficient q_{jkb} can be worsened to $\tilde{q}_{jkb}$ for each job j. So $\tilde{z}_{jkb}$ identify the edges with stressed costs.

For y and $\tilde{z}$ feasible, the maximization of the worst-case probability can be expressed as

$$\begin{aligned}
&\max_{y}\min_{\tilde{z}}\ \psi^{-1}\left(\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}((1-\tilde{z}_{jkb})q_{jkb}+\tilde{z}_{jkb}\tilde{q}_{jkb})y_{jkb}\right)\\
&=\psi^{-1}\left(\min_{y}\max_{\tilde{z}}\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}((1-\tilde{z}_{jkb})q_{jkb}+\tilde{z}_{jkb}\tilde{q}_{jkb})y_{jkb}\right)\\
&=\psi^{-1}\left(\min_{y}\left(\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}q_{jkb}y_{jkb}+\max_{\tilde{z}}\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}\Delta_{jkb}\tilde{z}_{jkb}y_{jkb}\right)\right),
\end{aligned}$$

where we used in the first equality that ψ and thus also ψ^{-1} is decreasing. In the second equality, we grouped the variables depending on $\tilde{z}$ and used the definition of $\Delta_{jkb}=\tilde{q}_{jkb}-q_{jkb}$.

Let us now take a look at the inner maximization problem for fixed y's feasible

$$\begin{aligned}
\max_{\tilde{z}}\ &\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}(\Delta_{jkb}y_{jkb})\tilde{z}_{jkb}\\
\text{s.t. }&\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}\tilde{z}_{jkb}=\Gamma\\
&0\leq\tilde{z}_{jkb}\leq 1,\quad \{j,k\}\in\bar{E}_b,\quad b\in\mathcal{B}
\end{aligned}$$

where these constraints depend only on $\tilde{z}$. Although $\tilde{z}$'s are binary, the maximization forces even variables from $[0,1]$ to be 1 in case it is not 0. Using the linear programming duality to the maximization problem, we get

$$\begin{aligned}
\min_{p,\,\theta}\ &\Gamma\theta+\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}p_{jkb}\\
\text{s.t. }&\theta+p_{jkb}\geq\Delta_{jkb}y_{jkb},\quad \{j,k\}\in\bar{E}_b,\quad b\in\mathcal{B} &\text{(4a)}\\
&p_{jkb}\geq 0,\quad \{j,k\}\in\bar{E}_b,\quad b\in\mathcal{B}. &\text{(4b)}
\end{aligned}$$

When we use $(a)^+=\max\{a,0\}$, we can write the dual problem as follows

$$\min_{\theta}\ \Gamma\theta+\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}(\Delta_{jkb}y_{jkb}-\theta)^+.$$

So the whole problem can be rewritten as

$$\begin{aligned}
\min_{y,\,\theta}\ &\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}q_{jkb}y_{jkb}+\Gamma\theta+\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}(\Delta_{jkb}y_{jkb}-\theta)^+\\
\text{s.t. }&(2a)-(2e).
\end{aligned}$$

And when we multiply the last term in the objective function by $1=\frac{\sum_{b\in\mathcal{B}}|\bar{E}_b|}{\sum_{b\in\mathcal{B}}|\bar{E}_b|}$, to get it in the form

$$\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}q_{jkb}y_{jkb}+\Gamma\left(\theta+\frac{\sum_{b\in\mathcal{B}}|\bar{E}_b|}{\Gamma}\sum_{b\in\mathcal{B}}\sum_{\{j,k\}\in\bar{E}_b}\frac{1}{\sum_{b\in\mathcal{B}}|\bar{E}_b|}(\Delta_{jkb}y_{jkb}-\theta)^+\right)$$

we may realize that applying CVaR, we get the stated reformulation, because we set $\alpha=1-\Gamma/\sum_{b\in\mathcal{B}}|\bar{E}_b|$ and thus $\frac{1}{\Gamma}\sum_{b\in\mathcal{B}}|\bar{E}_b|=\frac{1}{1-\alpha}$. □

3 Numerical Study

We will present two ways of problem solution and illustrate it on some generated problem instances. We do not have any real data of jobs and machines to schedule, so we had to simulate the data. We will now describe the procedure of simulation.

3.1 Problem Instance Generation

Firstly, we set the number of machine classes B and determine the number of machines C. We consider only the situation where every machine class has the same number of machines. For each of the machines, we generate K jobs. This means, we have a total of $J = C \cdot K$ jobs. The job generation on one machine is based on exponential distribution. We assume that the spaces between jobs and the job lengths have exponential distribution with parameters λ_1 and λ_2 respectively (parameters represent the expected value).

The parameter A representing number of job classes is then set as $A = B + 1$. We assume that every machine class can process two different job types and there is an overlay. For each machine class $b \in \mathcal{B}$ the pair of job types it can process is b and $b + 1$. We divide the job types equally between all jobs generated on a given machine class. It is done in a way that we assign the first job the first possible type, the second job is assigned the second type and so on.

This way, we obtain an instance of the problem which has a feasible schedule. For optimization, we increase the number of machines of each class by one to be sure that more than one feasible solution exists. Thus, the final number of machines is $C_+ = C + B$.

Then, we set the probability that a job is not delayed. This probability may depend on the machine type. This leads to a notation p_b for the probability that some job has no delay on machine of type $b \in \mathcal{B}$. The interpretation of this setting can be, that some machine classes are more reliable than others. If there is a delay, its length is assumed to have exponential distribution with parameter λ_3. The cumulative distribution function of a delay of job assigned to machine class b is then $F_b(x) = p_b + (1 - p_b)(1 - e^{-\lambda_3 x})$, where $x \geq 0$ similarly as in [1]. We consider $\psi(x) = -\log x$ as the copula generator function. That gives an independence of the delays. We use $\tilde{p}_j$ which is a random number simulated from uniform distribution on the interval $[0, 0.15]$. And the stressed probability is $\tilde{p}_{jb} = p_b - \tilde{p}_j$.

For the expected value of spaces and job lengths we chose $\lambda_1 = \lambda_2 = 5$. For the delays it was set $\lambda_3 = 2$. We considered the non-stressed probability of no delay to be 0.95 for the first machine type and decreasing by 0.05 for the following machine type, so that the B-th machine class has the probability of no delay $0.95 - 0.05 \cdot (B - 1)$.

3.2 Problem Solution with MIP Solver

Thanks to the duality, the problem (1) can be solved as

$$\min_{y,\, p,\, \theta} \left(\sum_{b \in \mathcal{B}} \sum_{\{j,k\} \in \bar{E}_b} q_{jkb} y_{jkb} + \Gamma\theta + \sum_{b \in \mathcal{B}} \sum_{\{j,k\} \in \bar{E}_b} p_{jkb} \right)$$

subject to constraints (2a) - (2e) and (4a) - (4b). We used Gurobi MIP solver [4], we had access to the academic licence. The solver was accessed through PuLP package in Python. For all the calculations, we used Python 3.8.12 on a computer with Intel(R) Core(TM) i5-4200U CPU @ 2.30 GHz, 16 GB RAM and 4 logical processors.

3.3 Golden-section Search

The problem can be also viewed as a two stage problem and decomposed in the way that the first-stage problem is

$$\min_{\theta} \Gamma\theta + \varphi(\theta) \text{ s.t. } \theta \in \left[0, \max_{\{j,k\} \in E_b,\, b \in \mathcal{B}} \Delta_{jkb} \right]$$

and the second-stage problem is

$$\varphi(\theta) = \min_{y} \left(\sum_{b \in \mathcal{B}} \sum_{\{j,k\} \in \bar{E}_b} q_{jkb} y_{jkb} + \sum_{b \in \mathcal{B}} \sum_{\{j,k\} \in \bar{E}_b} (\Delta_{jkb} y_{jkb} - \theta)^+ \right) \text{ s.t. } (2a) - (2e).$$

We will use the trick suggested by [2], for $y_{jkb} \in \{0,1\}$ and $\theta \in \left[0, \max_{\{j,k\}\in E_b,\, b\in\mathcal{B}} \Delta_{jkb}\right]$ it holds that $(\Delta_{jkb} y_{jkb} - \theta)^+ = (\Delta_{jkb} - \theta)^+ y_{jkb}$. So in fact the second-stage problem is simplified as follows

$$\varphi(\theta) = \min_{y} \left(\sum_{b\in\mathcal{B}} \sum_{\{j,k\}\in\bar{E}_b} (q_{jkb} + (\Delta_{jkb} - \theta)^+) y_{jkb} \right) \text{ s.t. } (2a) - (2e).$$

For solving this two-stage problem we will use golden-section search algorithm according to [2], which was also implemented on the robust FIS problem for homogeneous machines in [1] and we implemented it analogically.

3.4 Numerical Results

We chose three possible problem size settings and generated ten problem instances of each. Then we considered three possible number of jobs that have stressed delays Γ. The instances were identical for them, the only difference was in Γ. It was floored 10, 25, 40 % of all jobs. In the following table 1 we present the average computational times of the solver and the golden-section search (GSS) algorithm as well as the average number of iterations of the GSS algorithm over all ten generated problem instances. We can also find the average number of all edges in the

Size	$C_+ = 8,\ B = 2,\ J = 60$			$C_+ = 12,\ B = 3,\ J = 90$			$C_+ = 16,\ B = 4,\ J = 120$		
Γ	6	15	24	9	22	36	12	30	48
$\sum_{b\in\mathcal{B}} \lvert E_b \rvert$		2029.2			3779.7			5506.5	
$\sum_{b\in\mathcal{B}} \lvert \bar{E}_b \rvert$		1847.2			3476.7			5082.5	
Time of MIP solver	1.248	1.402	1.782	2.529	10.839	81.356	7.638	91.686	434.894*
Time of GSS	12.880	11.953	10.587	23.385	20.953	22.581	32.271	32.449	33.021
No. of iter. of GSS	24.9	24.9	24.9	25.0	25.0	25.0	25.0	25.0	25.0

Table 1 Average results over 10 instances, time in seconds.

* 4 out of 10 instances ended on the time limit of 600 second, their relative gaps were 1.7 − 3.2 %.

problem instances. As we can see, the average computational time of the GSS algorithm went from values around 10 seconds to values around 33 seconds. The average number of iterations was around 25 in all possible parameter settings. In case of the MIP solver, the computational time was much shorter for the smallest problem instances compared to the GSS algorithm in all possible settings of Γ. For the middle-sized instance it was better for 9 and 22 jobs with stressed delays. In case of 36 jobs with stressed delays, the computational time of the solver grew significantly. And for the greatest problem instance, the time of solver was better only for 12 jobs with stressed costs. When we stressed delays of 30 jobs, the average computational time was already three times longer than for the GSS algorithm. When stressing costs of 48 jobs, 4 out of 10 instances ended on the time limit of 600 seconds in case of the MIP solver. Their relative gaps were 1.7 − 3.2 %. And it significantly influenced the average solver computational time, which was considerably greater than for any other size setting with value over 7 minutes.

4 Conclusion

We derived a possible reformulation of a distributionally robust FIS problem which uses a risk measure CVaR. We suggested and implemented two ways to solve the problem in hands and compared them on couple of simulated problem instances of multiple sizes. The approach using a golden-section search algorithm lead to stable computational times in contrast to the Gurobi MIP solver, which ended on the time limit for some of the greatest problem instances.

References

[1] Branda, M. (2018). Distributionally robust fixed interval scheduling on parallel identical machines under uncertain finishing times. *Computers & Operations Research*, 98, 231-239. https://doi.org/10.1016/j.cor.2018.05.025.

[2] Bertsimas, D. & Sim, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming*, 98 (1), 49-71. https://doi.org/10.1007/s10107-003-0396-4.

[3] Rockafellar, R. & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26 (7), 1443-1471. https://doi.org/10.1016/S0378-4266(02)00271-6.

[4] Gurobi Optimization, LLC (2023). *Gurobi Optimizer Reference Manual*. [Online]. Available at: https://www.gurobi.com [cited 2023-06-11].

System Dynamics Modelling Scenarios for Economic-ecological System of the Aral Sea

Mira Mauleshova[1]

Abstract. The paper is dedicated to the implementation of System Dynamics modelling in the domain of the Shrinking Aral Sea's catastrophe. The problem of irrational Water Management in the agricultural sector and its adverse impact on the environment in the Republic of Uzbekistan is modelled by means of System Dynamics techniques.

In the framework of stated problem, the Stock and Flow diagram is constructed in order to outline the causality between demographical structure, overwhelming water usage in agroeconomic activities and its influences on environmental subsystems such as immersive diminishment of the Aral Sea's water volume as well as degradation of its water quality from the standpoint of its salinity.

In the paper the model construction is defined, subsequently, model and its parameters are validated. Furthermore, the paper includes the baseline as well as what-if scenarios of the key variables, namely, the volume of water flowing to the Aral Sea, the volume of the Aral Sea and its salinity.

Keywords: Aral Sea, Computer simulation, System Dynamics, Water Management, scenarios

JEL Classification: C44, C63, Q15
AMS Classification: 90B90, 93C15

1 Introduction

In the 1960s, the sown area of the cotton industry in Central Asia was expanded twofold by the massive irrigation canals construction project. The consequence of the irrigation processes is a sharp decrease in the volume of the Amu Darya River, which inflows into the Aral Sea. This fact led to the dramatic shrinkage of the Aral Sea's volume, consequently increasing its salinity [10]. Only 10 % of the original area of the Aral Sea is currently left. Moreover, the lake's origin was divided into 4 separate parts [8]. The unsuccessful attempts of groundwater usage to workaround the catastrophe created another issue of its limit exceeding [14].

Within the paper, the focus is on the part of the Aral Sea, located in the Republic of Uzbekistan, due to the fact that it still tackles the issue of the Aral Sea's water volume shortage. The crucial point is that even though both rivers, the Amu Darya as well as the Syr Darya, are used to satisfy population needs, only the Amu Darya River inflows into the part of the Aral Sea located in the Republic of Uzbekistan [6], [17]. The Republic of Uzbekistan uses 91% of consumed water for agricultural purposes. Moreover, 4.2 million ha out of 5.2 million ha of agricultural land are irrigated [4], [11]. These facts lead to the enormous water consumption, nearly $2*10^6$ liters per capita, which enclose Uzbekistan among the highest consumers of water per capita around the globe, taking it to 4^{th} place [5].

In the paper, we focus on the scenarios of the economic-ecological system of the Aral Sea. The construction of the System Dynamics model was introduced in the previous paper "Dynamics of the Economic-Ecological System of Aral Sea" [12]. The model connects socio-economic and environmental subsystems. Therefore, the decision of the past leads not only to the shrinkage of the Aral Sea but also to the side effects, namely, sandstorms and health issues of the inhabitants.

2 Material and Methods

The modeling process is based on the steps introduced in [18]. Once the problem of ineffective Water Management was articulated, the dynamics hypothesis was formulated using a causal loop diagram, emphasizing key feedback loops. The Stock and Flow diagram was built to point out the physical structure. Figure 1 represents a simplified

[1] CZU Prague, Department of Systems Engineering, Kamýcká 129, 165 00 Prague, mauleshova@pef.czu.cz

version of the Stock and Flow diagram, which initially contains 101 symbols, namely, 5 levels, 45 auxiliaries, and constants/parameters 26.

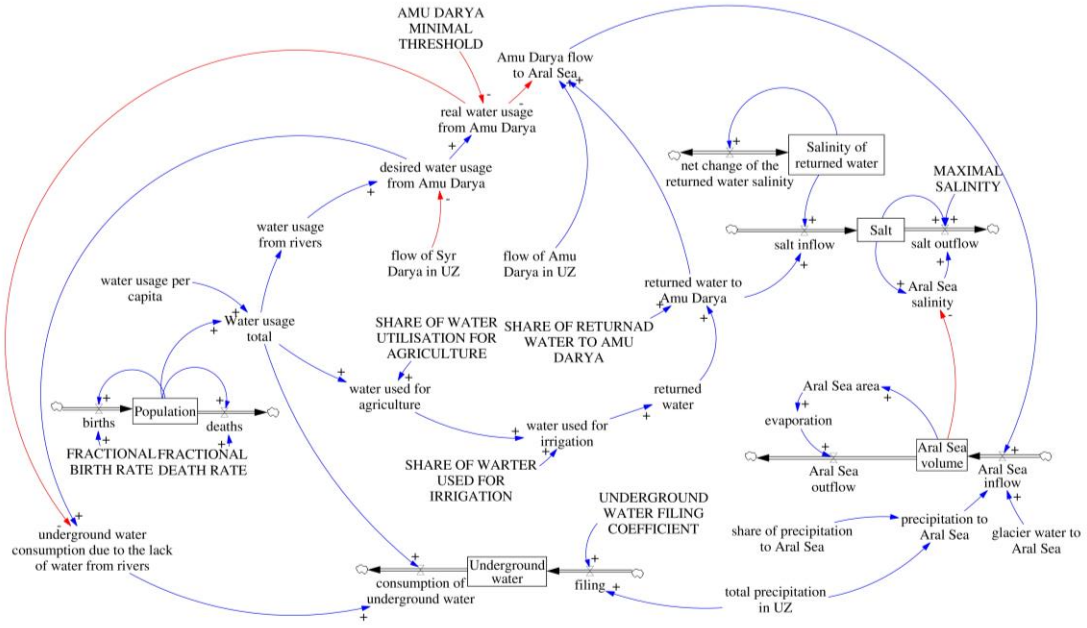


Figure 1 the Stock and Flow diagram of the economic-ecological system of the Aral Sea

Rectangles represent the stock/level variables. In and out flows/rates are represented by the picture of the pipe. For simulation purposes, the stock variable is defined as follows [18]:

$$s_T \int_{T_0}^{T} (i_t - o_t)\mathrm{d}t + s_{T_0}, \tag{1}$$

Where s_T is stock, i_t is all inflows, o_t are all outflows, T_0 is the initial time, T is the current time, and t is any time between T and T_0.

All causal links have defined polarity. The positive causal links' polarity of the link from x to y is defined [18]:

$$\frac{\delta y}{\delta x} > 0, \tag{2}$$

and similarly, the negative polarity [18]:

$$\frac{\delta y}{\delta x} < 0. \tag{3}$$

The official data within the time series 1998-2017 were used in the model depending on up-to-date data availability. To be precise, the statistics of water consumption, precipitation, surface water as well as groundwater are based on the core database of FAO AQUASTAT [6]. Moreover, the details on precipitation close to the Aral Sea were taken into account [9]. From the source named CAWater [3], the statistics were obtained not only about actual water consumption but also about drainage as well as water returned to the Amu Darya River. The population statistics were taken from United Nations statistics [20]. Data related to the Pamirs glacier, from which melting water flows to the Aral Sea, is acquired from the following source [2].

Powell optimisation was implemented to calibrate missing parameters of salt in the lake [15]. The model was verified by the dimensional analysis test and sensitivity analysis [18].

3 Results and Discussion

The model's evaluation and sensitivity analysis were detailly described in the previous paper [12]. To rebrief, the selected indicators, *MAPE* and R^2, of the critical variables indicated acceptable accuracy of the selected simulated data compared to the real ones. Sensitivity analysis didn't detect any specific leverage point. The model is evaluated as accurate; therefore, scenarios of the selected variables are simulated within the present paper.

The Scenario of Change in the Birth Rate

First and foremost, there are improper mental models in society that the expansion of irrigated land and, consequently drastically increased water demand is justified by the need to ensure rations to the growing population. Therefore, the first selected scenario simulates an impact caused by the birth rate change to the Aral Sea volume. The original average value of the birth rate variable is 0.022438381 dmnl/year. The simulated change deals with *a decrease of the variable's value by 5%, 10%, 20%*, and *an increase by 10% in the birth rate*. Additionally, *a stable population* is simulated, whereby the death rate equals the original value of the birth rate.

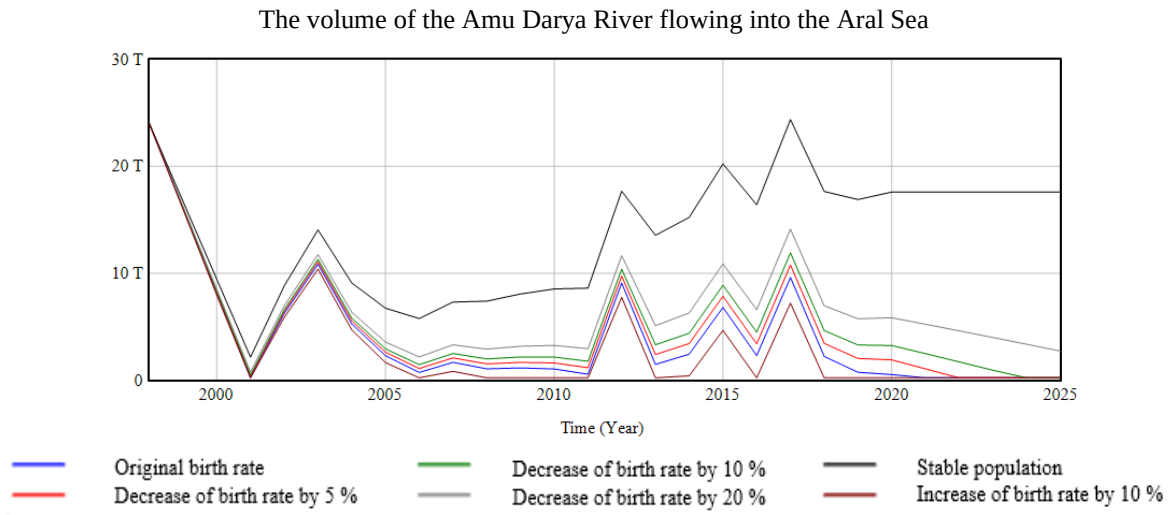


Figure 2 the Amu Darya River flowing into the Aral Sea within the scenarios of change in the birth rate

According to Figure 2 within *a stable population* scenario, water flowing from the Amu Darya River reaches the Aral Sea in a significantly larger volume. Logically, with the more significant drop in the birth rate, the greater volume of water goes from the river to the Aral Sea. In case of *a 5% and 10% drop*, there is a need for additional water sources, namely, groundwater pumping. In case of *a 20% drop in the birth rate*, even though the simulated water values of the Amu Darya River decrease, the function does not reach the minimum value. This fact indicates that there would be enough water in the river as a source for irrigation within the scenario of *a 20% drop in the birth rate.*

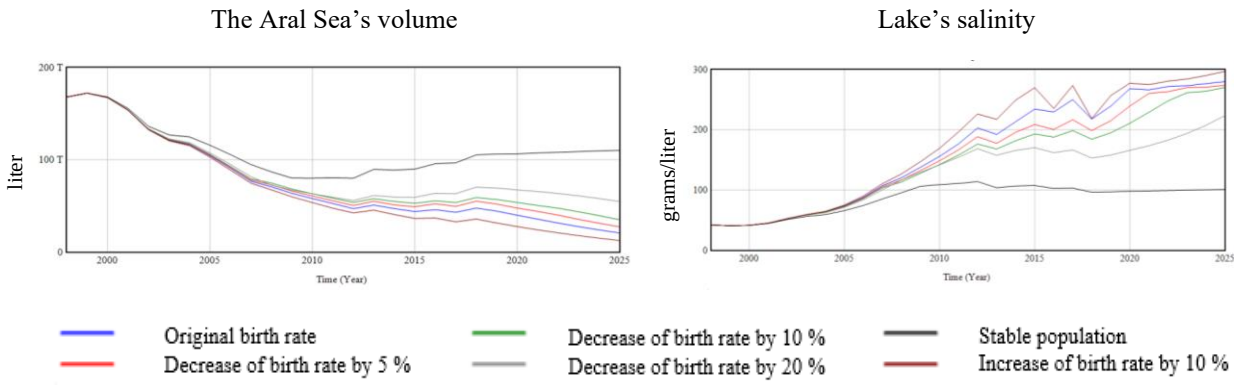


Figure 3 The Aral Sea's volume and salinity – scenarios of change in the birth rate

Figure 3 signifies the lake's volume and salinity dynamics within the simulated scenarios. It is worth highlighting that *reduction in the birth rate by 20%* has a slight positive effect not only on the volume of the Aral Sea, which accounts for 39.39%, but also on salinity. Only in the case of *a stable population*, the Aral Sea volume is not

decreasing. Moreover, the lake's salinity within *a stable population* does not exceed the barrier of 150 grams per liter, but this scenario is less likely. It is observable that the lake's salinity approaches the maximum salinity value under all the simulated scenarios except for *20% of the birth rate reduction* and *stable population scenarios.*

The Scenario of the Share of Demand Satisfied by Surface Water

The demand for total water consumption is met by means of 2 resources. The former is surface water, which fulfills 87% of water demand on average. The rest is satisfied by the latter resource, namely, groundwater.
The scenario deals with the simulation when the variable *the share of demand satisfied by surface water* in Uzbekistan was reduced to the *levels 60%, 70% from 2010*, due to the fact that in this year, the Eastern part of the South Aral Sea completely dried up [1].

The lower water volume is taken from the river since *the share of demand satisfied from surface water drops to 60% and 70% from 2010*. Based on Figure 4, within the scenarios, higher water volume is getting to the Aral Sea. It is also seen that thanks to the increase in the volume of the Aral Sea under the scenario of *the share of demand satisfied from surface water reduction to 60% since 2010*, the salinity of the lake has been below 100 grams per liter since 2015 and has not exceeded the mentioned value over time. In case of scenario of *the share of demand satisfied from surface water reduction to 70% since 2010*, the salinity has the same trend as in case of *reduction of variable to 60% since 2010.*

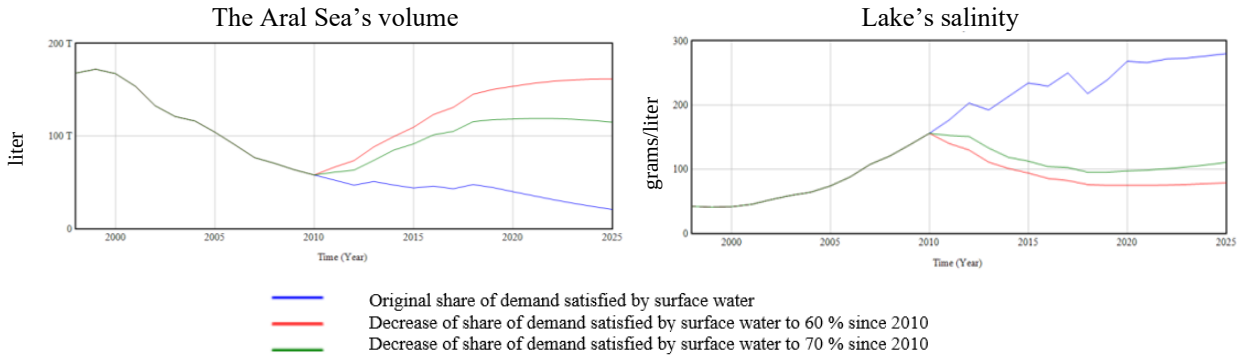


Figure 4 The Aral Sea's volume and salinity – scenarios of share of demand satisfied by surface water

On the other hand, due to the decrease in *the share of demand satisfied from surface water at the level of 60% and 70%*, there is a need to fulfill the total demand by groundwater pumping. The decision to decrease *the share of demand satisfied by surface water* negatively affected not only the pumping of underground water, shown in Figure 5, but also the condition of underground water in Uzbekistan. The consequences are shown in right part of Figure 5. To be precise, since 2010, the variable *Aggregate balance of groundwater* has been sharply decreasing, it leads to a zero value in the case of the scenario where *the share of demand satisfied from surface water decreased to 70%* in 2014 and further to negative values. In the scenario *of the share of demand satisfied from surface water reduction to 60%*, the same situation occurs, but 3 years earlier.

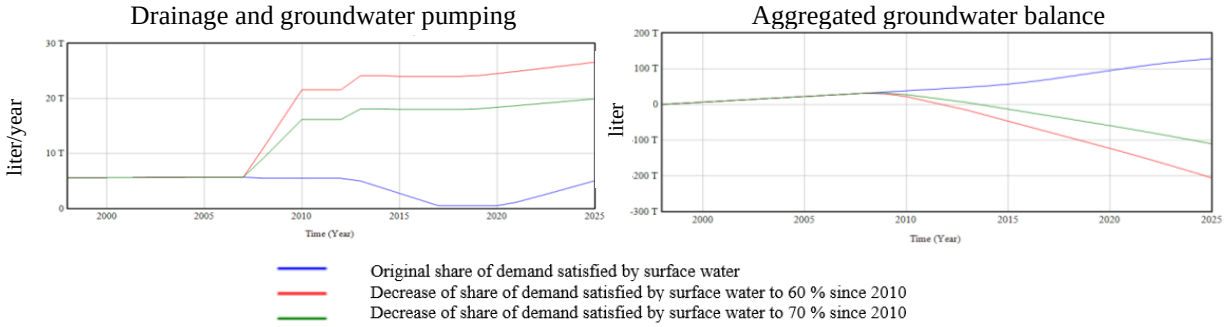


Figure 5 Drainage, groundwater and balance

The Scenario of the Share of Water Used in Agricultural Activities

Water consumption for agricultural purposes accounts 91% of total water consumption in Uzbekistan. Based on the scenarios, the variable *share of water used in agricultural activitie*s was reduced to the levels 60% and 70% during the whole simulated period. The same proceeded since 2018 as alternative scenarios.

Figure 6 illustrates the evolution of the Aral Sea volume. The original scenario shows a negative trend in the development of the mentioned variable, although a period of negligible growth in the volume of the lake can be noticed. Within the scenario of *the share of water used for agricultural activity reduction to the level of 60%, 70%,* a larger volume of water from the Amu Darya River enters the Aral Sea. Therefore, the lake does not experience any ecological collapse. Figure 6 reveals the relation between salinity and the volume of the lake, which can be interpreted in such a way that the larger the volume of the Aral Sea, the lower the salinity of the lake, i.e., a negative linkage.

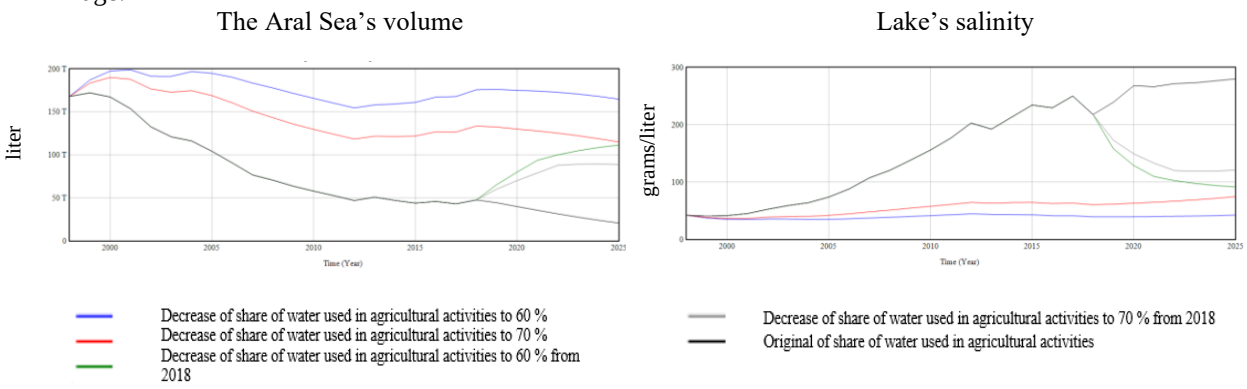


Figure 6 The Aral Sea's volume and salinity – scenarios of water usage for agricultural activities

It is worth mentioning that the changes were tested in historical time series. Additionally, even though the toxic environment due to sandstorms causes serious health issues for inhabitants, the model does not include a loop of the quality of the environment's improvement that could probably influence people's health and, thus life expectancy.

4 Conclusion

The article focuses on the scenarios of the economic-ecological system of the Aral Sea. Based on the above analyses, it can be concluded that the Republic of Uzbekistan experiences water scarcity and utilizes an excessive amount of water per capita. At first glance, water scarcity is caused by the growing population, which boosts the water demand. However, the scenario of *the change in the birth rate* points out that even though the birth rate had dropped by 20 %, the volume of the Aral Sea would still have decreased, but at slower tempo. However, the reality is that the population of the country is growing rapidly. The scenario *of the share of demand satisfied by surface water* indicates that the workaround of decreasing surface water usage by the expansion of groundwater pumping is just another way of symptom solving. This quick solution creates another issue from a long-term perspective, namely, the depletion of underground water, which cannot be considered an adequate solution. According to the scenario of *change in the share of water used in agricultural activities,* the tipping point is when *the share of water used for agricultural activity is reduced to 60%, 70%* of the total water needed in 2018. The scenarios show an ultimate change in the trend of the development of the volume of the Aral Sea from 2018, which means that if the approach to Water Management in the agricultural sector had been changed, it would have been possible to save the lake.

Yesterday's "solutions" are the cause of today's problems [16]. There is a delay between short-term benefits and long-term harmful consequences [18]. The main reasons for the difference between them are structures of feedback loops, delays, and non-linearity, which originate the complexity of the structure. The results of the structure's complexity "suddenly" pop up in the long-term period. Therefore, the focus needs to be on the long-term view [13], [16]. However, most of the time it seems easier to use solutions that give immediate effect without forward thinking.

Based on system thinking, there are always limits to growth. No physical entity can grow indefinitely. If the human population does not impose its own limits to keep growth within the capacity of the supporting environment, then the environment will impose the limits [13]. Once the structure and the cause of the problem are comprehended,

there is a need for the response in the proper time. Once a limit is exceeded, there is almost no way to go back [19]. The catch is that the later the change is implemented, there is less impact of the change furthermore the higher the costs of change [7]. Dominance is gained by a positive loop, which tends to move the system into imbalance, causing a significant difference in the state [19].

Acknowledgements

The research was supported by the project 2021B0003 of the Internal Grant Agency of the Faculty of Economics and Management CZU Prague (IGA FEM).

References

[1] Britanica, The Editors of Encyclopaedia. (2021). *Aral Sea.* [Online]. Available at: https://www.britannica.com/place/Aral-Sea [cited 2022-02-06].
[2] Brun, F., Berthier, E., Wagnon, P., Kääb, A. & Treichler, D. (2017). *A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000-2016*. Nature Geoscience, 10, 668–673. https://doi.org/10.1038/ngeo2999.
[3] CAWater-Info, Portal of Knowledge for Water and Environmental Issues in Central Asia. (2022). *Database of the Aral Sea*. [Online]. Available at: http://cawater-info.net/aral/data/index_e.htm [cited 2022-02-06].
[4] Central Asian Bureau for Analytical Reporting. (2020). *Uzbekistan's Water Sector: Environmental and Managerial Issues*. [Online]. Available at: https://cabar.asia/en/uzbekistan-s-water-sector-environmental-and-managerial-issues [cited 2022-02-26].
[5] EURASIANET. (2022). *Central Asian States Are World's Leading Water Wasters.* [Online]. Available at: https://eurasianet.org/central-asian-states-are-worlds-leading-water-wasters [cited 2022-02-20].
[6] Food and Agriculture Organization of the United Nations. (2021). *AQUASTAT Core Database. Food and Agriculture Organization of the United Nations*. [Online]. Available at: https://www.fao.org/aquastat/en/databases/maindatabase/ [cited 2022-02-05].
[7] Harrell, C., Ghosh, B. K. & Bowden, R. (2012). *Simulation using ProModel* (3rd edition), New York, NY: McGraw-Hill.
[8] Chen, D.-H. (2018). *The country that brought a sea back to life*. [Online]. Available at: https://www.bbc.com/future/article/20180719-how-kazakhstan-brought-the-aral-sea-back-to-life [cited 2022-02-04].
[9] Krapivin, V., Mkrtchyan, F. & Rochon, G. (2019). *Hydrological Model for Sustainable Development in the Aral Sea Region.* Hydrology, 6, 91. https://doi.org/10.3390/hydrology6040091.
[10] Kulturologia.ru. (2021). *Как в СССР загубили Аральское море и зачем хотели повернуть вспять сибирские реки*. [Online]. Available at: https://kulturologia.ru/blogs/210921/51138/ [cited 2022-02-06].
[11] Lioubimtseva, E. (2014). *Impact of Climate Change on the Aral Sea and its Basin*. In: Micklin P., Aladin N. & Plotnikov I. (Eds.) The Aral Sea. Berlin: Springer. 405-427. https://doi.org/10.1007/978-3-642-02356-9_17.
[12] Mauleshova, M., Krejčí, I. (2022). Dynamics of the Economic-Ecological System of Aral Sea. *In 40th International Conference on Mathematical Methods in Economics, Conference proceedings* Jihlava. Jihlava: College of Polytechnics Jihlava. pp. 229-234.
[13] Meadows, D. H. (2008). *Thinking in systems: a primer*. Wright, D. (Ed.) White River Junction, Vt.: Chelsea Green Pub.
[14] Micklin, P. (2010). *The past, present, and future Aral Sea.* Lakes & Reservoirs: Research and Management. 15, 193–213. https://doi.org/10.1111/j.1440-1770.2010.00437.x.
[15] 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.
[16] Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization*, New York: Doubleday/Currency.
[17] Shmakova, U. (2018*). Национальный отчет по управлению возвратными водами в республике Узбекистан.* [Online]. Available at: http://riverbp.net/Отчет+по+управлению+возвратными+водами+в+Республике+Узбек истан.pdf [cited 2022-02-04].
[18] Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World.* Boston: Irwin/McGraw-Hill.
[19] Šusta, M. (2015). *Průvodce systémovým myšlením*. Praha: Proverbs, 136 p., ISBN 978-80-260-7602-5.
[20] United Nations. (2019). *World population prospects*. [Online]. Available at: https://population.un.org/wpp/ [cited 2022-01-08].

Strong Robustness of Convex and Concave Monge Matrices in Max-min Algebra

Monika Molnárová[1]

Abstract. Strong robustness of both convex and concave Monge matrices over max-min algebra is studied. Strong robustness of a matrix is connected with the greatest solution of the eigenproblem, i.e. the greatest eigenvector corresponding to the matrix. In this paper important properties of threshold digraphs of convex and concave Monge matrices in regard to matrix strong robustness are pointed out. Equivalent conditions for period equal to one of a strongly connected threshold digraph for a convex Monge matrix are presented. The period of a strongly connected threshold digraph for a concave Monge matrix is shown to be equal one. Moreover, the strong connectivity can be verified effectively by checking the existence of cycles of every pair of consecutive nodes. Necessary and sufficient conditions for convex Monge matrix and concave Monge matrix to be strongly robust were proved. Using obtained results effective algorithms for checking strong robustness in both cases are described.

Keywords: max-min algebra, convex Monge matrix, concave Monge matrix, strong robustness

JEL Classification: C02
AMS Classification: 08A72, 90B35, 90C47

1 Introduction

To solve optimization problems in many diverse areas ([3], [10], [11]) using the concept of extremal algebras is a frequent method to model discrete dynamic systems. We have studied Monge matrices, their structural properties and algorithms solving many problems related to Monge matrices in [1], [4], [5] and [6]. Results concerning robustness and strong robustness of matrices were presented in [2], [4], [7], [8] and [9].

The aim of this paper is to present results related to the structure of the threshold digraphs of convex and concave Monge matrices over max-min algebra, to prove equivalent conditions for convex and concave Monge matrices to be strongly robust and introduce more effective algorithms for checking the strong robustness of the considered types of matrices with fixed data.

We briefly outline the content and main results of the paper. Section 2 provides the necessary preliminaries on max-min algebra, the greatest eigenvector and the property of strong robustness of a max-min matrix are defined and necessary and sufficient conditions for strong robustness in general case of a max-min matrix are recalled. Finally, the notion of a convex and concave Monge matrix is introduced. In Section 3, we prove equivalent conditions for the period of a strongly connected threshold digraph to be equal to one in Theorem 3, a necessary and sufficient condition for strong robustness of a convex Monge matrix in Theorem 4 and a more effective algorithm for checking the strong robustness in Theorem 5. In Section 4, we prove an equivalent condition for the acyclicity of a threshold digraph in Theorem 8, an equivalent condition for the strong connectivity of a threshold digraph in Theorem 10, a necessary and sufficient condition for strong robustness of a concave Monge matrix in Theorem 11 and a more effective algorithm for checking the strong robustness in Theorem 12.

2 Background of the Problem

The max-min algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus$, $\otimes$. The least element in B will be denoted by O, the greatest one by I. By $\mathbb{N}$ we denote the set of all natural numbers. For a given natural $n \in \mathbb{N}$, we use the notation N for the set of all smaller or equal positive natural numbers, i.e., $N = \{1, 2, \ldots, n\}$.

For any $m, n \in \mathbb{N}$, $B(m,n)$ denotes the set of all matrices of type $m \times n$ 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.

[1] Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Monika.Molnarova@tuke.sk

A *digraph* is a pair $G = (V,E)$, where V, the so–called vertex set, is a finite set, and E, the so–called edge set, is a subset of $V \times V$. A path in the digraph $G = (V,E)$ is a sequence of vertices $p = (i_1, \dots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1, \dots, k$. The number k is the length of the path p and is denoted by $\ell(p)$. If $i_1 = i_{k+1}$, then p is called a cycle. For a given matrix $A \in B(n,n)$ the symbol $G(A) = (N,E)$ stands for the complete, edge-weighted digraph associated with A, i.e., the vertex set of $G(A)$ is N, and the capacity of any edge $(i,j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A,h)$ is the digraph $G = (N,E')$ with the vertex set N and the edge set $E' = \{(i,j);\ i,j \in N,\ a_{ij} \geq h\}$. By a *strongly connected component* of a digraph $G(A,h) = (N,E)$ we mean a subdigraph $\mathcal{K} = (N_\mathcal{K}, E_\mathcal{K})$ generated by a non-empty subset $N_\mathcal{K} \subseteq N$ such that any two distinct vertices $i,j \in N_\mathcal{K}$ are contained in a common cycle, $E_\mathcal{K} = E \cap (N_\mathcal{K} \times N_\mathcal{K})$ and $N_\mathcal{K}$ is the maximal subset with this property. A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. For any non-trivial strongly connected component $\mathcal{K}$ is the *period* of $\mathcal{K}$ defined as $\operatorname{per}\mathcal{K} = \gcd\{\ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per}\mathcal{K} = 1$. By $\mathrm{SCC}^\star(G)$ we denote the set of all non-trivial strongly connected components of G.

Obviously, it is enough to consider thresholds $h \in H = \{a_{ij};\ i,j \in N\}$ to get all threshold digraphs corresponding to a matrix A.

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 $V(A,\lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.

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*.

The greatest $\lambda-$eigenvector corresponding to the matrix $A = (a_{ij}) \in B(n,n)$ and $\lambda \in B$ is defined as

$$x^\star(A,\lambda) = \bigoplus_{x \in V(A,\lambda)} x, \tag{1}$$

It was proved that for $\lambda = I$ the greatest eigenvector exists and an iterative procedure for computing it was introduced in [2].

Similarly as in [7] we can define the strong robustness for every matrix $A = (a_{ij}) \in B(n,n)$ using the following values

$$c_i(A) = \bigoplus_{j \in N} a_{ij} \qquad c(A) = \bigotimes_{i \in N} c_i(A) \tag{2}$$

$$c^+(A) = \min\{a_{ij};\ a_{ij} > c(A)\} \tag{3}$$

Definition 2. Let $A = (a_{ij}) \in B(n,n)$, $\lambda \in B$. Let $c^\star(A) = (c(A), c(A), \dots, c(A)) \in B(n)$. Let $M(A) = \{x \in B(n);\ x < c^\star(A)\}$. Let $T^\star(A,\lambda) = \{x \in B(n);\ x^\star(A,\lambda) \in O(A,x)\}$. The matrix A is called *strongly $\lambda-$ robust* if

$$T^\star(A,\lambda) = B(n) \setminus M(A). \tag{4}$$

A strongly λ–robust matrix with $\lambda = I$ is called *a strongly robust matrix*.

A sufficient and necessary condition for strong robustness of a max-min matrix was proved and an $O(n^3)$ algorithm for verifying the strong robustness was introduced in [7].

Theorem 1. [7] *Let $A \in B(n,n)$. Then A is a strongly robust matrix if and only if $G(A,c^+(A))$ is acyclic and $G(A,c(A))$ is a strongly connected digraph with period equal to one.*

Definition 3. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is *a convex Monge matrix* (*concave Monge matrix*) if and only if

$$a_{ij} \otimes a_{kl} \leq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l$$

$$(\ a_{ij} \otimes a_{kl} \geq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l\ ).$$

3 Strong Robustness of Convex Monge Matrices

In this section we present results crucial for formulating a necessary and sufficient condition for convex Monge matrices to be strongly robust. Consequently we present a more effective algorithm for verifying the strong robustness of a convex Monge matrix.

Theorem 2. [5] *Let $A \in B(n,n)$ be a convex Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A,h))$. Let c be a cycle of length $\ell(c) \geq 3$ in $\mathcal{K}$. Then c can split in $\mathcal{K}$ into finite number of cycles of length one or two.*

The above result allows to check whether the period of a strongly connected digraph equals to one without computing the period of the digraph.

Theorem 3. *Let $A \in B(n,n)$ be a convex Monge matrix. Let $G(A,h)$ be strongly connected for $h \in H$. Then the period of the digraph $G(A,h)$ equals to one if and only if there is a loop in $G(A,h)$.*

Proof. Let $h \in B$. Let the period of the threshold digraph $G(A,h)$ equal to 1. Since $G(A,h)$ consists of one strongly connected component the period is the greatest common divisor of the lengths of all cycles in $G(A,h)$. Moreover by Theorem 2 every cycle can split into the cycles of length one and two. Consequently there is a loop in $G(A,h)$.
For the converse implication let us assume that there is a loop in $G(A,h)$. Since the digraph $G(A,h)$ is strongly connected the result follows by definition of the period of a strongly connected component. □

Now, using the obtained results we can reformulate the necessary and sufficient condition from Theorem 1 for a convex Monge matrix to be strongly robust.

Theorem 4. *Let $A \in B(n,n)$ be a convex Monge matrix. Then A is a strongly robust matrix if and only if $G(A,c^+(A))$ is acyclic and $G(A,c(A))$ is a strongly connected digraph with a loop.*

Proof. The theorem is a consequence of Theorem 1 and Theorem 3. □

The computational complexity of the algorithm for checking the strong robustness for a max-min matrix described in [7] is $O(n^3)$. We can improve the computational complexity for convex Monge matrices using the above results. Namely, instead of computing the period of a digraph we verify the existence of at least one loop in the considered digraph.

Theorem 5. *There is an algorithm with computational complexity $O(n^2)$ for verifying the strong robustness of a convex Monge matrix.*

Proof. To determine the computational complexity let us recall the well-known $O(n^2)$ algorithm for checking the acyclicity and strong connectivity of a digraph. First, we compute $c(A)$ and $c^+(A)$ in $O(n^2)$ time. Second, we check the acyclicity of $G(A,c^+(A))$. Third, we check the strong connectivity of $G(A,c(A))$. Finally, we verify by condition $a_{ii} \geq c(A)$ the existence of a loop in $G(A,c(A))$ for some $i \in N$ in $O(n)$ time. Hence the total computational complexity of the algorithm is $O(n^2)$. □

Example 1. Let us check the strong robustness of the convex Monge matrix $A \in B(6,6)$ for $B = [0,3]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 & 2 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

According to Theorem 5 we start with computation of the value $c(A)$ finding by formulas (2) the row maxima for every row and afterwards the minimal element of them, i.e. $c(A) = \min\{3, 3, 2, 1, 1, 1\} = 1$. We proceed with determining $c^+(A)$ as the minimal element of the set of all elements greater than $c(A)$ using formula (3), i.e. $c^+(A) = \min\{2, 3\} = 2$. The corresponding threshold digraph $G(A,c^+(A)) = G(A,2)$ is acyclic (see Figure 1), while the threshold digraph $G(A,c(A)) = G(A,1)$ is strongly connected and contains a loop (see Figure 2). Hence the considered convex Monge matrix is strongly robust.

$G(A, c^+(A)) = G(A, 2)$

Figure 1 Acyclic threshold digraph $G(A, c^+(A))$ of the convex Monge matrix

$G(A, c(A)) = G(A, 1)$

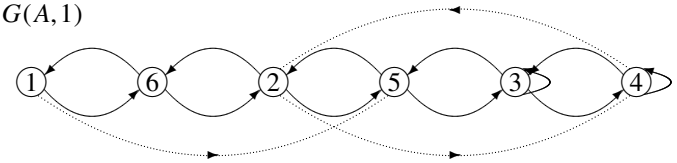


Figure 2 Strongly connected threshold digraph $G(A, c(A))$ of the convex Monge matrix

4 Strong Robustness of Concave Monge Matrices

In this section we present results crucial for formulating a necessary and sufficient condition for concave Monge matrices to be strongly robust. Consequently we present a more effective algorithm for verifying the strong robustness of a concave Monge matrix.

In contrast to the convex Monge matrices, where every cycle can split into the cycles of length two and one, in case of concave Monge matrices the decomposition of a cycle contains only cycles of length one.

Theorem 6. [6] *Let $A \in B(n,n)$ be a concave Monge matrix. Let c be a cycle in $G(A,h)$ for $h \in H$. Then c can split in $\mathcal{K}$ into finite number of cycles of length one.*

Theorem 7. [6] *Let $A \in B(n,n)$ be a concave Monge matrix. Let c be a cycle in $G(A,h)$ for $h \in H$. Then there is a loop on every node of the cycle c.*

Corollary 1. [6] Let $A \in B(n,n)$ be a concave Monge matrix. Let $\mathcal{K} \in \mathrm{SCC}^\star(G(A,h))$ for $h \in H$. Then there is a loop on every node in $\mathcal{K}$.

We can reformulate the previous corollary using the definition of the period of a strongly connected component.

Corollary 2. Let $A \in B(n,n)$ be a concave Monge matrix. Let $\mathcal{K} \in \mathrm{SCC}^\star(G(A,h))$ for $h \in H$. Then the period of $\mathcal{K}$ is equal to one.

The above results are useful to prove the following necessary and sufficient condition for a threshold digraph of a concave Monge matrix to be acyclic.

Theorem 8. *Let $A \in B(n,n)$ be a concave Monge matrix. The digraph $G(A,h)$ for $h \in H$ is acyclic if and only if $G(A,h)$ contains no loop.*

Proof. Let $h \in H$. Let the digraph $G(A,h)$ contain no loop. Let us assume there is a cycle c in $G(A,h)$. Using Theorem 7 there is a loop on every node of the cycle c in $G(A,h)$, what is a contradiction. Consequently the considered digraph is acyclic.
The converse implication follows by definition of acyclicity of a digraph. □

The nodes of a strongly connected component of a threshold digraph corresponding to a concave Monge matrix are numbered by a sequence of consecutive natural numbers ([6]). The following theorem helps to prove a necessary and sufficient condition for a threshold digraph of a concave Monge matrix to be strongly connected.

Theorem 9. [6] *Let $A \in B(n,n)$ be a concave Monge matrix. Let i and $i+1$ be two nodes in a strongly connected component $\mathcal{K}$ of $G(A,h)$ for $h \in H$. Then $\mathcal{K}$ contains the cycle $(i, i+1, i)$.*

Theorem 10. *Let $A \in B(n,n)$ be a concave Monge matrix. The digraph $G(A,h)$ for $h \in H$ is strongly connected if and only if $G(A,h)$ contains the cycle $(i,i+1,i)$ for $i = 1, 2, \ldots, n-1$.*

Proof. Let $h \in H$. Let the digraph $G(A,h)$ be strongly connected. Using Theorem 9 there is the cycle $(i,i+1,i)$ in $G(A,h)$ for every $i = 1, 2, \ldots, n-1$.
The converse implication follows by definition of strong connectivity of a digraph. □

Now, using the obtained results we can reformulate the necessary and sufficient condition from Theorem 1 for a concave Monge matrix to be strongly robust.

Theorem 11. *Let $A \in B(n,n)$ be a concave Monge matrix. Then A is a strongly robust matrix if and only if $G(A,c^+(A))$ contains no loop and $G(A,c(A))$ is a strongly connected digraph.*

Proof. The theorem is a consequence of Theorem 1, Theorem 8, Theorem 10 and Corollary 2. □

Similarly as in case of convex Monge matrices we can improve the computational complexity of the algorithm for checking the strong robustness for a max-min matrix. The obtained results for concave Monge matrices essentially decrease the time necessary for checking the strong connectivity or acyclicity of a digraph. By checking only the diagonal elements we can decide whether the digraph is acyclic. The existence of a common cycle of length two for each pair of consecutive nodes guaranties the strong connectivity of the digraph. In addition, it is not necessary to compute the period of a digraph, since the strong connectivity implies the existence of loops in the considered digraph and consequently its period equals one.

Theorem 12. *There is an algorithm for verifying the strong robustness of a concave Monge matrix*
(i) with computational complexity $O(n^2)$,
(ii) with computational complexity $O(n)$ if $c(a)$ and $c^+(A)$ are given.

Proof. The computation of $c(A)$ and $c^+(A)$ in $O(n^2)$ time dominates the total computational complexity of the algorithm. The acyclicity of $G(A,c^+(A))$ due to Theorem 8 can be verified by condition $a_{ii} < c^+(A)$ for every $i \in N$ in $O(n)$ time. The strong connectivity of $G(A,c(A))$ can be verified by checking the existence of cycles of length two for each pair of consecutive nodes due to Theorem 10 in $O(n)$ arithmetic operations. Hence the total computational complexity of the algorithm is $O(n^2)$, without computation of $c(A)$ and $c^+(A)$ is the computational complexity $O(n)$. □

Example 2. Let us check the strong robustness of the concave Monge matrix $A \in B(8,8)$ for $B = [0,3]$

$$A = \begin{pmatrix} 1 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

According to Theorem 12 we start with computation of the value $c(A)$ by formulas (2) finding the row maxima for every row and afterwards the minimal element of them, i.e. $c(A) = \min\{3, 2, 1, 1, 2, 2, 3, 1\} = 1$. We proceed with determining $c^+(A)$ as the minimal element of the set of all elements greater than $c(A)$ using formula (3), i.e. $c^+(A) = \min\{2, 3\} = 2$. For checking the acyclicity of the corresponding threshold digraph $G(A,c^+(A)) = G(A,2)$ it is enough to prove that there is no loop in the digraph, i.e. to verify the condition $a_{ii} < 2$ for every $i \in N$ (see Figure 3). We use Theorem 10 to prove the strong connectivity of the threshold digraph $G(A,c(A)) = G(A,1)$ by checking the existence of cycles $(i,i+1,i)$ for each $i = 1, 2, \ldots, n-1$ (see Figure 4). Hence the considered concave Monge matrix is strongly robust.

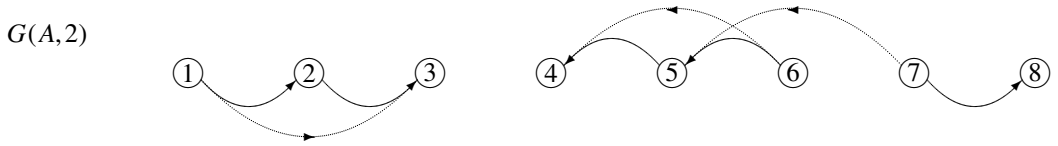


Figure 3 Acyclic threshold digraph $G(A, c^+(A))$ of the concave Monge matrix

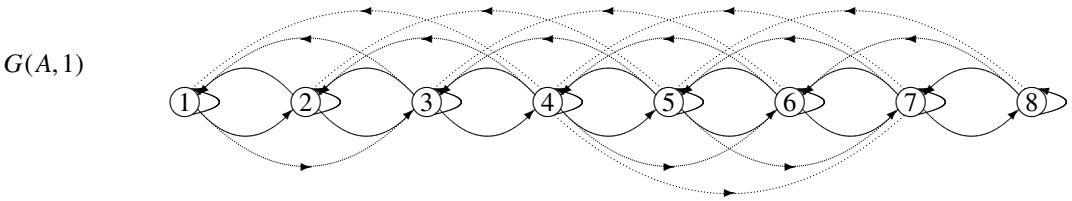


Figure 4 Strongly connected threshold digraph $G(A, c(A))$ of the concave Monge matrix

5 Conclusion

We have studied properties of threshold digraphs with respect to the greatest eigenvector of convex and concave Monge matrices over max-min algebra in this paper. Obtained results were used to prove necessary and sufficient conditions for strong robustness and to derive more effective algorithms for checking the strong robustness of considered classes of matrices with fixed data. Obtained results can be used investigating robustness, strong robustness and further types of robustness connected with diverse types of eigenvectors ([3]) for matrices and interval matrices.

The matrix strong robustness is connected with the problem of finding the greatest eigenvector of a given matrix. Looking for the greatest solution of the eigenproblem can represent the following economic application ([9]). A system supports web users buying products $p_1, p_2, \ldots, p_n$. Let A represent preferences, i.e. a_{ij} describes the preference p_i to p_j. What are the maximum levels of interest for the products which are not influenced by A? The question leads to the problem of finding the greatest eigenvector of the given matrix A.

References

[1] Burkard, R. E., Klinz, B. & Rudolf, R. (1996). Perspectives of Monge properties in optimization, *DAM*, Volume 70, 95–161.

[2] Cechlárová, K. (1997). Efficient computation of the greatest eigenvector in fuzzy algebra, *Tatra Mt. Math. publ.*, Volume 12, 73–79.

[3] Gavalec, M., Ramík, J. & Zimmermann, K. (2015). *Decision making and optimization*. Lecture Notes in Economics and Mathematical Systems 677, Springer.

[4] Hireš, M., Molnárová, M. & Drotár, P. (2020). Robustness of Interval Monge Matrices in Fuzzy Algebra, *Mathematics*, Volume 4, 1–16.

[5] Molnárová, M. (2020). Periodicity of convex and concave Monge matrices in max-min algebra, In: *Proceedings of* 38^{th} *Int. Conference Mathematical Methods in Economics 2020*, Brno, 377–382.

[6] Molnárová, M. (2021). Structure of the threshold digraphs of convex and concave Monge matrices in max-min algebra, In: *Proceedings of* 39^{th} *Int. Conference Mathematical Methods in Economics 2021*, Praha, 331–336.

[7] Plavka, J. & Szabó, P. (2012). On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume 160, 640–647.

[8] Plavka, J. (2011). On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices, *DAM*, Volume 159 Issue 5, 381–388.

[9] Sanchez, E. (1978). Resolution of eigen fuzzy sets equation, *Fuzzy Sets and Systems*, Volume 1, 69–74.

[10] De Schutter, B., van den Boom, T., Xu, J. & Farahani, S. S. (2020). Analysis and control of max-plus linear discrete-event systems: An introduction. *Discrete Event Dyn. Syst.*, Volume 30, 25–54.

[11] Zimmermann, H. J. (2011). *Fuzzy Set Theory And Its Applications*. Springer Science and Business Media, Berlin, Germany.

On the von Neumann Regularity of Max-min Matrices

Helena Myšková[1]

Abstract. The 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, 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. Many practical situations can be described using max-min matrix equations. Several properties of matrices in max-min algebra have been studied. One of them is von Neumann regularity, which means that there is a matrix X such that $A \otimes X \otimes A = A$.

Keywords: max-min algebra, interval matrix, von Neumann regularity, von Neumann X-regularity

JEL classification: C02
AMS classification: 15A18, 15A80, 65G30

1 Introduction

1.1 Background of the Problem

The behaviour 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. Discrete dynamic systems and related algebraic structures were studied using max-min matrix operations in [2], [4], [13].

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. A generalization of systems of linear equations are matrix equations of the form $A \otimes X \otimes C = B$, where A, B and C are given matrices of suitable sizes and X is an unknown matrix, which have been studied in [3],[7]. In the case that $A = B = C$, the solvability of $A \otimes X \otimes A = A$ means that A is von Neumann regular. In max-min theory, there is three-way correspondence between von Neumann regularity, idempotency and projectivity, see [6]. In this paper, we shall require the existence of the solution of $A \otimes X \otimes A = A$ with entries in given intervals of possible values. In this way, several types of the von Neumann regularity can be defined. We study the von Neumann regularity, possibly, universally, AE and EA von Neumann regularity according to $\boldsymbol{X}$. Polynomially testable equivalent conditions are given for each of them. The results are illustrated by numerical examples. Other types of regularity in max-min algebra have been studied in [1], [4], [12]. In classical algebra, AE regularity of interval matrices discussed in [5].

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovakia, helena.myskova@tuke.sk

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 capacity of the road from P_i to Q_j (from T_l to D_k). If there is no road from P_i to Q_j (from T_l to 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 $\min\{a_{ij}, x_{jl}, c_{lk}\}$. Let b_{ik} denotes the number of passengers travelling from place P_i to destination D_k. Our task is to choose the appropriate capacities x_{jl} for any $j \in N = \{1,\ 2, \ldots,\ n\}$, and for any $l \in S = \{1, 2, \ldots, s\}$ such that the maximum capacity of the road from P_i to D_k is equal to a given number b_{ik} for any $i \in M = \{1, 2, \ldots, m\}$ and for any $k \in R = \{1, 2, \ldots, r\}$, i. e.,

$$\max_{j \in N,\, l \in S} \min\{a_{ij}, x_{jl}, c_{lk}\} = \ b_{ik}. \tag{1}$$

1.2 Preliminaries

By max-min (fuzzy) algebra, we understand a triplet $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I}$ is a linearly ordered set, $\oplus = \max$, $\otimes = \min$ are binary operations on $\mathcal{I}$. The notation $\mathcal{I}(m, n)$ ($\mathcal{I}(n)$) denotes the set of all matrices (vectors) of given dimensions over $\mathcal{I}$. Operations $\oplus$, $\otimes$ are extended to matrices and vectors in a standard way. The linear ordering on $\mathcal{I}$ also induces partial orderings on $\mathcal{I}(m, n)$ and $\mathcal{I}(n)$. Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in the classical algebra. We will 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 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 B, $D \in \mathcal{I}(n, s)$ the implication inequalities $A \leq C$, $B \leq D$ imply $A \otimes B \leq C \otimes D$.

Using max-min algebra, equation (1) can be written in the form of matrix equation

$$A \otimes X \otimes C = B. \tag{2}$$

where $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(s, r)$. We are looking for $X \in \mathcal{I}(n, s)$ such that

2 Von Neumann Regularity

Definition 1. A matrix $A \in \mathcal{I}(n, n)$ is von Neumann regular if there is $X \in \mathcal{I}(n, n)$ such that

$$A \otimes X \otimes A = A. \tag{3}$$

To give equivalent conditions for the von Neumann regularity, let us recall some known results about the solvability of (2).

Denote by $X^*(A, B, C) = \big(x^*_{jl}(A, B, C)\big)$ the matrix defined as follows

$$x^*_{jl}(A, B, C) = \min_{k \in R}\{x^*_j(A \otimes c_{lk}, B_k)\}, \tag{4}$$

where

$$x^*_j(A, b) = \min_{i \in M}\{b_i : \ a_{ij} > b_i\} \tag{5}$$

for any $j \in N$, where $\min \emptyset = I$. We shall call the matrix $X^*(A, B, C)$ a *principal matrix solution* of (2). The following theorem expresses the properties of $X^*(A, B, C)$ and gives the necessary and sufficient condition for the solvability of (2).

Theorem 1. [3] *Let $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(s, r)$.*

(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* $X^*(A,B,C)$ *is its solution.*

Remark 1. Equality (4) can be written in the form

$$X^*(A,B,C) = (X_1^*(A,B,C), X_2^*(A,B,C), \ldots, X_s^*(A,B,C)),$$

where

$$X_l^*(A,B,C) = \min_{k \in R} x^*(A \otimes c_{lk}, B_k) \tag{6}$$

If $A = B = C \in \mathcal{I}(n,n)$, we obtain

$$x_{jl}^*(A,A,A) = \min_{k \in N}\{x_j^*(A \otimes a_{lk}, A_k)\} = \min_{k \in N} \min_{i \in N}\{a_{ik} : a_{ij} \otimes a_{lk} > a_{ik}\}.$$

In the following, we shall use notation $X^*(A)$ instead of $X^*(A,A,A)$. We have

$$x_{jl}^*(A) = \min_{i,k \in N}\{a_{ik} : a_{ij} \otimes a_{lk} > a_{ik}\}. \tag{7}$$

Theorem 2. *A matrix* $A \in \mathcal{I}(n,n)$ *is von Neumann regular if and only if* $A \otimes X^*(A) \otimes A = A$.

Proof. The proof follows directly from Theorem 1 (iii). □

Next, we will require that the solution of (3) is not arbitrary, but that it takes on values from the given interval of possible values. Similarly to [3, 8, 9], for a given $\underline{X}, \overline{X} \in \mathcal{I}(n,n)$, $\underline{X} \leq \overline{X}$ define an given interval

$$\boldsymbol{X} = [\underline{X}, \overline{X}] = \{X \in \mathcal{I}(n,n) : \underline{X} \leq X \leq \overline{X}\}.$$

Definition 2. A matrix $A \in \mathcal{I}(n,n)$ is

1. possibly von Neumann regular according to $\boldsymbol{X}$ if there is $X \in \boldsymbol{X}$ such that $A \otimes X \otimes A = A$,
2. universally von Neumann regular according to $\boldsymbol{X}$ if for any $X \in \boldsymbol{X}$ the equality $A \otimes X \otimes A = A$ holds.

Let us define $X^*(A, \overline{X}) = \min\{X^*(A), \overline{X}\}$.

Theorem 3. *Let* $A \in \mathcal{I}(n,n)$ *and* $\boldsymbol{X} \subseteq \mathcal{I}(n,n)$ *be given. Then* A *is possibly von Neumann regular according to* $\boldsymbol{X}$ *if and only if* $X^*(A, \overline{X}) \geq \underline{X}$ *and* $A \otimes X^*(A, \overline{X}) \otimes A = A$.

Proof. Suppose that $X^*(A, \overline{X}) \geq \underline{X}$ and $A \otimes X^*(A, \overline{X}) \otimes A = A$. Since $X^*(A, \overline{X}) \leq \overline{X}$, we have $X^*(A, \overline{X}) \in \boldsymbol{X}$. The equality $A \otimes X^*(A, \overline{X}) \otimes A = A$, implies that there exists $X \in \boldsymbol{X}$, namely $X = X^*(A, \overline{X})$ such that $A \otimes X \otimes A = A$. Hence A is possibly von Neumann.

For the converse implication suppose that either

$$X^*(A, \overline{X}) \ngeq \underline{X} \quad \text{or} \quad A \otimes X^*(A, \overline{X}) \otimes A \neq A.$$

If there exists $j, l \in N$ such that $x_{jl}^*(A, \overline{X}) < \underline{x}_{jl}$, then $x_{jl}^*(A) < \underline{x}_{jl}$. According to Theorem 1 (i), $x_{jl} < \underline{x}_{jl}$ for each $X \in \boldsymbol{X}$ such that $A \otimes X \otimes A = A$, so there is no solution of (3) in $\boldsymbol{X}$. Then A is not possibly von Neumann regular according to $\boldsymbol{X}$.

If $A \otimes X^*(A, \overline{X}) \otimes A \neq A$, then according to Theorem 1 (ii) we obtain that there exist $i, k \in N$ such that $[A \otimes X^*(A, \overline{X}) \otimes A]_{ik} < a_{ik}$. We shall prove that $[A \otimes X \otimes A]_{ik} < a_{ik}$ for each $X \in \boldsymbol{X}$. Since $[A \otimes X^*(A, \overline{X}) \otimes A]_{ik} = \max_{j,l \in N} \min\{a_{ij}, x_{jl}^*(A, \overline{X}), a_{lk}\}$, we have $\min\{a_{ij}, x_{jl}^*(A, \overline{X}), a_{lk}\} < a_{ik}$ for any $j, l \in N$. Let $j, l \in N$ be arbitrary, but fixed. If $x_{jl}^*(A, \overline{X}) = x_{jl}^*(A)$, then Theorem 1 (i) implies that $\min\{a_{ij}, x_{jl}, a_{lk}\} < a_{ik}$ for each $X \in \mathcal{I}(n,n)$. If $x_{jl}^*(A, \overline{X}) = \overline{x}_{jl}$, then $\min\{a_{ij}, x_{jl}, a_{lk}\} \leq \min\{a_{ij}, \overline{x}_{jl}, a_{lk}\} < a_{ik}$ for each $x \in \boldsymbol{X}$. Then $[A \otimes X \otimes A]_{ik} < a_{ik}$ for each $X \in \boldsymbol{X}$, so A is not possibly von Neumann regular according to $\boldsymbol{X}$. □

Theorem 4. *Let $A \in \mathcal{I}(n,n)$ and $\boldsymbol{X} \subseteq \mathcal{I}(n,n)$ be given. A matrix $A \in \mathcal{I}(n,n)$ is universally von Neumann regular according to $\boldsymbol{X}$ if and only if $X^*(A) \geq \overline{X}$ and $A \otimes \underline{X} \otimes A = A$.*

Proof. If $X^*(A) \geq \overline{X}$ and $A \otimes \underline{X} \otimes A = A$, then for each $X \in \boldsymbol{X}$ we have

$$A = A \otimes \underline{X} \otimes A \leq X \otimes A \leq A \otimes \overline{X} \otimes A \leq A \otimes X^*(A) \otimes A \leq A.$$

Hence $A \otimes X \otimes A = A$ for each $A \in \boldsymbol{X}$, so A is universally von Neumann.

For the converse implication suppose that either $X^*(A) \ngeq \overline{X}$ and $A \otimes \underline{X} \otimes A \neq A$.

If $X^*(A) \ngeq \overline{X}$, then there exists $j, l \in N$ such that $x^*_{jl}(A) < \overline{x}_{jl}$. Then for each $X \in \boldsymbol{X}$ such that $x_{jl} > x^*_{jl}(A)$ we obtain $A \otimes X \otimes A \neq A$, so A is not universally von Neumann regular.

If $A \otimes \underline{X} \otimes A \neq A$, then for $X = \underline{X}$ the equality (3) does not hold, so A is not universally von Neumann regular. □

2.1 EA and AE von Neumann Regularity

In the previous section, we required either the existence of $X \in \boldsymbol{X}$ that (3) is satisfied or the validity of (3) for each $X \in \boldsymbol{X}$. In the following, we will consider the case that some elements of the matrix X must be taken into account for each value of the interval, and for some of them it is enough to be considered for at least one value.

The AE and EA concepts for the regularity have been studied, see [5], [10],[11]. Suppose that each interval $[\underline{x}_{ij}, \overline{x}_{ij}]$ is associated either with the universal, or with the existential quantifier. Then we can split the interval matrix as $\boldsymbol{X} = \boldsymbol{X}^{\forall} \oplus \boldsymbol{X}^{\exists}$, where $\boldsymbol{X}^{\forall}$ is the interval matrix comprising universally quantified coefficients and $\boldsymbol{X}^{\exists}$ concerns existentially quantified coefficients. Thereafter denote by $\tilde{N}^{\exists} \subseteq N \times N$ and $\tilde{N}^{\forall} \subseteq N \times N$ the corresponding sets of indices. In other words, $\underline{x}^{\exists}_{ij} = \overline{x}^{\exists}_{ij} = O$ for each couple $(i,j) \in \tilde{N}^{\forall}$ and $\underline{x}^{\forall}_{ij} = \overline{x}^{\forall}_{ij} = O$ for each couple $(i,j) \in \tilde{N}^{\exists}$.

Definition 3. Let $A \in \mathcal{I}(n,n)$, $\boldsymbol{X} \subseteq \mathcal{I}(n,n)$ and $\tilde{N}^{\exists}$ and $\tilde{N}^{\forall}$ be given. We say that A is

1. *EA-von Neumann regular* according to $\boldsymbol{X}$ if

$$(\exists X^{\exists} \in \boldsymbol{X}^{\exists})(\forall X^{\forall} \in \boldsymbol{X}^{\forall}) A \otimes (X^{\exists} \oplus X^{\forall}) \otimes A = A$$

2. *AE-von Neumann regular* according to $\boldsymbol{X}$ if

$$(\forall X^{\forall} \in \boldsymbol{X}^{\forall})(\exists X^{\exists} \in \boldsymbol{X}^{\exists}) A \otimes (X^{\exists} \oplus X^{\forall}) \otimes A = A$$

With respect to the assignment of the existential and general quantifier to the individual components of the interval vector $\boldsymbol{X}$, we can also write

$$X^*(A) = X^{*\forall}(A) \oplus X^{*\exists}(A), \quad X^*(A, \overline{X}) = X^{*\forall}(A, \overline{X}) \oplus X^{*\exists}(A, \overline{X}).$$

Theorem 5. *Let $A \in \mathcal{I}(n,n)$, $\boldsymbol{X} \subseteq \mathcal{I}(n,n)$ and $\tilde{N}^{\exists}$ and $\tilde{N}^{\forall}$ be given. Then A is EA-von Neumann regular according to $\boldsymbol{X}$ if and only if*

$$X^{*\exists}(A) \geq \underline{X}^{\exists}, \ X^{*\forall}(A) \geq \overline{X}^{\forall} \quad \text{and} \quad A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A = A. \tag{8}$$

Proof. Suppose that A is EA-von Neumann regular according to $\boldsymbol{X}$, i. e., there exists $\tilde{X^{\exists}} \in \boldsymbol{X}^{\exists}$ such that $A \otimes (\tilde{X^{\exists}} \oplus X^{\forall}) \otimes A = A$ for each $X^{\forall} \in \boldsymbol{X}^{\forall}$. It follows that A is universally von Neumann regular according to $\tilde{\boldsymbol{X}} = [\tilde{X^{\exists}}, \tilde{X^{\exists}}] \oplus [\underline{X}^{\forall}, \overline{X}^{\forall}]$. By Theorem 4 we obtain $X^{*\exists}(A) \geq \tilde{X^{\exists}} \geq \underline{X}^{\exists}$, $X^{*\forall}(A) \geq \overline{X}^{\forall}$ and $A \otimes (\tilde{X^{\exists}} \oplus \underline{X}^{\forall}) \otimes A = A$. Moreover, the inequality $\tilde{X^{\exists}} \leq X^{*\exists}(A, \overline{X})$ follows from $\tilde{X^{\exists}} \in \boldsymbol{X}^{\exists}$. Then for $X^{\exists} = X^{*\exists}(A, \overline{X})$ we obtain

$$A = A \otimes (\tilde{X^{\exists}} \oplus \underline{X}^{\forall}) \otimes A \leq A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A \leq$$

$$A \otimes (X^{*\exists}(A) \oplus \underline{X}^{\forall}) \otimes A \leq A \otimes X^{*}(A) \otimes A \leq A.$$

Hence $A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A = A$.

For the converse implication suppose that (8) is satisfied. Then $X^{*\exists}(A, \overline{X}) \in \boldsymbol{X}^{\exists}$ and for each $X^{\forall} \in \boldsymbol{X}^{\forall}$ we obtain

$$A = A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A \leq A \otimes (X^{*\exists}(A, \overline{X}) \oplus X^{\forall}) \otimes A \leq$$

$$A \otimes (X^{*\exists}(A, \overline{X}) \oplus \overline{X}^{\forall}) \otimes A \leq A \otimes (X^{*\exists}(A) \oplus X^{*\forall}(A)) \otimes A \leq A \otimes X^{*}(A) \otimes A \leq A$$

Then there exists $X^{\exists} \in \boldsymbol{X}^{\exists}$, namely $X^{\exists} = X^{*\exists}(A, \overline{X})$ such that $A \otimes (X^{*\exists}(A, \overline{X}) \oplus X^{\forall}) = A$ for each $X^{\forall} \in \boldsymbol{X}^{\forall}$. Hence A is EA-von Neumann regular according to $\boldsymbol{X}$. □

Theorem 6. *Let $A \in \mathcal{I}(n,n)$, $\boldsymbol{X} \subseteq \mathcal{I}(n,n)$ and $\tilde{N}^{\exists}$ and $\tilde{N}^{\forall}$ be given. Then A is AE-von Neumann regular according to $\boldsymbol{X}$ if and only if it is EA-von Neumann regular according to $\boldsymbol{X}$.*

Proof. Suppose that A is AE-von Neumann regular according to $\boldsymbol{X}$. First, the condition $X^{*\forall}(A) \geq \overline{X}^{\forall}$ trivially follows. Then for $X^{\forall} = \underline{X}^{\forall}$ there exists $\tilde{X^{\exists}} \in \boldsymbol{X}^{\exists}$ such that $A \otimes (\tilde{X^{\exists}} \oplus \underline{X}^{\forall}) \otimes A = A$. Then $\underline{X}^{\exists} \leq \tilde{X^{\exists}} \leq X^{*\exists}(A, \overline{X}) \leq X^{*\exists}(A)$. We obtain

$$A = A \otimes (\tilde{X^{\exists}} \oplus \underline{X}^{\forall}) \otimes A \leq A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A \leq A \otimes X^{*}(A) \otimes A \leq A.$$

Hence $A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A = A$, $X^{*\exists}(A) \geq \underline{X}^{\exists}$ and $X^{*\forall}(A) \geq \overline{X}^{\forall}$. According to Theorem 5, A is EA-von Neumann regular according to $\boldsymbol{X}$.

The converse implication is trivial. □

Example 2. Let $\mathcal{I} = [0, 10]$ and let $A = \begin{pmatrix} 5 & 1 & 5 \\ 10 & 1 & 4 \\ 6 & 9 & 8 \end{pmatrix}$. Decide whether

i) A is von Neumann regular;

ii) A is possibly von Neumann regular and universally von Neumann regular according to $\boldsymbol{X}$, where

$$\boldsymbol{X} = \begin{pmatrix} [1,3] & [10,10] & [0,1] \\ [4,8] & [2,5] & [7,9] \\ [2,7] & [3,6] & [0,1] \end{pmatrix};$$

iii) AE(EA)-von Neumann regular according to $\boldsymbol{X}$ if

$$N^{\exists} = \{(1,2), (2,3), (3,1), (3,2)\}, \quad N^{\forall} = N \times N - N^{\exists}.$$

Solution. i) We compute $X^{*}(A)$ by formula (4) and check whether the condition of Theorem 2 is satisfied. We obtain

$$X^{*}(A) = \begin{pmatrix} 4 & 10 & 1 \\ 10 & 6 & 10 \\ 10 & 6 & 1 \end{pmatrix}, \qquad A \otimes X^{*}(A) \otimes A = A.$$

Hence A is von Neumann regular.

ii) Universal regularity: $X^{*}(A) \geq \overline{X}$, but $A \otimes \underline{X} \otimes A \neq A$, so A is not universally von Neumann regular, according to Theorem 4.
Possible regularity: We have

$$X^{*}(A, \overline{X}) = \begin{pmatrix} 3 & 10 & 1 \\ 8 & 5 & 9 \\ 7 & 6 & 1 \end{pmatrix} \geq \underline{X}, \qquad A \otimes X^{*}(A, \overline{X}) \otimes A = A.$$

According to Theorem 3, the given matrix is possibly von Neumann regular.

iii) We check whether the conditions of Theorem 5 are fulfilled. We have $X^{*\exists}(A) \geq \underline{X}^{\exists}$, and $X^{*\forall}(A) \geq \overline{X}^{\forall}$. It remains to check whether $A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A = A$. We obtain

$$A \otimes (X^{*\exists}(A, \overline{X}) \oplus \underline{X}^{\forall}) \otimes A = A \otimes \begin{pmatrix} 1 & 10 & 0 \\ 4 & 2 & 9 \\ 7 & 6 & 0 \end{pmatrix} \otimes A = A,$$

so the given matrix is EA-von Neumann regular. According to Theorem 6, it is AE-von Neumann regular, too.

2.2 Conclusion

In this paper, we dealt with the von Neumann regularity in max-min algebra, which is the property that is related to the projectivity of max-min (tropical) polytopes. On the other hand, it is related to the solvability of a special type of matrix equation. Using the solvability of max-min matrix equations is a useful tool for describing real situation in economics and industry. In Example 1, the values a_{ij}, x_{jl} and c_{lk} represent the capacities of corresponding connections. In economics, those values may represent for example the financial costs for the production or transporting of some products. Another possibility is that a_{ij} represents a measure of the preference of the property P_i of some object before the property Q_j, similarly x_{jk} represent a measure of the preference of the property Q_j before the property D_k. In practice, the capacities (financial costs, preferences) are from certain intervals rather than fixed numbers. Therefore, it would be appropriate if the matrix A were also interval. The study of von Neumann regularity of interval matrix is our main goal for the future.

References

[1] Cechlárová, K. (1995). Unique solvability of max-min fuzzy equations and strong regularity of matrices over fuzzy algebra. *Fuzzy Sets and Systems*, 75, 165–177.

[2] Cuninghame-Green, R. A. (1979). *Minimax Algebra*. Lecture notes in Economics and Mathematical systems.

[3] Draženská, E. & Myšková, H. (2017). Interval fuzzy matrix equations. *Kybernetika*, 53 (1), 99–112.

[4] Gavalec, M. & Plavka, J. (2003). Strong regularity of matrices in general max–min algebra. *Lin. Algebra Appl.*, 371, 241–254.

[5] Hladík M. (2018). AE regularity of interval matrices. *Journal of Linear Algebra*, 33, 137–146.

[6] Izhakian, Z., Johnson, M. & Kambites, M. (2016). Pure dimension and projectivity of tropical polytopes. *Advances in Mathematics*, 303 (5), 1236–1263.

[7] Myšková, H. (2016). Interval max-plus matrix equations. *Lin. Algebra Appl.*, 492 , 111–127.

[8] Myšková, H. (2005). Interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 403, 263–272.

[9] Myšková, H. & Plavka, J. (2013). X-robustness of interval circulant matrices in fuzzy algebra. *Lin. Algebra Appl.*, 438, 2757–2769.

[10] Myšková, H. & Plavka, J. (2020). AE and EA robustness of interval circulant matrices in max-min algebra. *Fuzzy Sets and Systems*, 384, 91–104.

[11] Myšková, H. & Plavka, J. (2021). AE and EA robustness of interval circulant matrices in max-product algebra. *Fuzzy Sets and Systems*, 410, 45–59.

[12] Myšková, H. & Plavka, J. (2023). Regularity of interval fuzzy matrices. *Fuzzy Sets and Systems*, 463, https://doi.org/10.1016/j.fss.2023.01.013

[13] Zimmermann, U. (1981). *Linear and combinatorial optimization in ordered algebraic structures*, North Holland, Amsterdam.

Effect of Factor Numbers in the Approximation of Returns on Portfolio Performance

David Neděla[1]

Abstract. Many strategies for setting the number of factors even simple or advanced have been implemented in a returns approximation process. Thus, in this contribution, the objective of this contribution is to analyse the effect of two simple strategies for determining an appropriate number of factors obtained from the principal component analysis (PCA) on the performance of the portfolio. Moreover, we also analyse the impact of the application of several dependency matrices to PCA. To approximate the return series, this study employs a nonparametric regression model based on conditional expectations and kernel estimator. The effect on the constructed portfolio is shown employing a portfolio model that maximizes the Sharpe ratio. Furthermore, different time periods of US stock market data are considered.

Keywords: multifactor analysis, nonparametric regression, performance measuring, principal component analysis

JEL Classification: C14, C38
AMS Classification: 62G08, 62H25

1 Introduction

Approximation of returns based on factors is an actual and important topic related to the portfolio selection process, [4], [6], [8]. Most of these works employ the factors obtained by applying principal component analysis (PCA). The question that arises from this issue is how to properly determine the number of factors. To this date, the literature has developed many advanced techniques, which are focusing on the procedure to determine the appropriate number of factors obtained from a particular multifactor model, see, among others, [1], [2], [3], [13]. In general, we can distinguish two simple strategies for setting the number of factors included, targeting either the number of factors or the explanatory power. Specifically, with the factor number strategy, we use a static (constant) number of factors in all applications. Otherwise, the strategy considering explanation power uses a dynamic number of factors in each application but leads to the required explanation of total portfolio variability. These strategies are easy-applicable to individual investors. However, another important question is how the number of factors included affects the performance of the constructed portfolio. These questions are one of the motivating factors of this work.

For this reason, this contribution aims to examine the relationship between the number of factors obtained from PCA used for return approximation and portfolio performance while capturing various market conditions. Moreover, to extend our analysis, we also apply the PCA to different dependency matrices, i.e., variance–covariance, Pearson and trend-dependent correlations. For this reason, this work represents a kind of sensitivity analysis. Since we consider the approximation of returns from different factors, we use a non-parametric regression including Ruppert and Wand (RW) estimation for this procedure, [6], [9].

The empirical part of this study is based on the US market dataset (components of S&P 500 index), where three different time periods are considered. This division is more appropriate for examining consequent influences and connections with their comments. From the results, we can deduce whether selected strategies and dependency matrices are suitable for particular market conditions. An important takeaway of this analysis is the advantageous properties of the trend correlation for PCA.

The rest of this contribution is divided as follows. In Section 2, the description of the methodology used is provided. Section 3 presents the dataset used, a description of an empirical procedure, results achieved and their discussion. The whole contribution is concluded in Section 4.

2 Methodology

In this section, we present the principle of multifactor analysis using PCA, the nonparametric regression approach based on Ruppert and Wand multivariate locally weighted least squares regression [9], and selected types of dependency measures used in portfolio management [8], [10].

[1] VSB – Technical University of Ostrava, Department of Finance, Sokolská třída 33, Ostrava, Czech Republic, david.nedela@vsb.cz

We denote $x = [x_1, x_2, \ldots, x_z]$ as a vector of asset weights and $r = [r_1, r_2, \ldots, r_z]$ as a vector of gross returns. Thus, $x'r$ represents the vector of portfolio returns.

In multifactor analysis, we can substitute the original z dependent series (e.g. of returns) $\{r_i\}_{i=1}^{z}$ with z new uncorrelated series denoted for example as $\{w_i\}_{i=1}^{n}$ employing e.g., the PCA method. In doing so, each r_i can be defined as a specific function of w_i series. Thus, if we assume parametric approach using the linear regression model, the i-th return vector r_i is derived only by selected k factors obtained from the PCA as follows:

$$r_i = \alpha_i + \sum_{j=1}^{k} \beta_{i,j} f_j + \varepsilon_i, \quad \text{for } i = 1, \ldots, z, \tag{1}$$

where α_i is the constant value, $\beta_{i,j}$ is the coefficient related to the factor f_j, and ε_i is the error part of the i-th asset estimation with mean equals 0.

Nonparametric Regression Model

However, in this study, we consider a alternative regression method that fits better the aspects of financial return series [6] or [7]. In particular, we apply the nonparametric regression method, which is characterized by including conditional expectations and a multivariate kernel estimator. Assuming the determination of the factors by performing PCA on individual correlation matrix, the formulation of the nonparametric approach is as follows:

$$r = E(r \mid F = f) + \epsilon = m(f) + \epsilon, \tag{2}$$

where E represents the expected value, $f = (f_1, \ldots, f_k)$ means the vector of k uncorrelated factors, and ϵ is the error of estimation. To determine the function $m(f)$, we use the locally weighted least squares regression estimator introduced by [9]. As stayed in [9], the main task related to the estimation of regression function $m(f)$ is to find the solution of the parameter a, which is solved by optimization task:

$$\min_{a,b} \sum_{i=1}^{T} [r_i - a - b^T (f_i - f)]^2 K_H (f_i - f), \tag{3}$$

where $K_H(\cdot)$ is a multivariate kernel estimator with an $s \times s$ symmetric positive definite matrix H depending on the sample size T, and f_i is the i-th observation of vector of factor f. For a more detailed discussion and application, see e.g. [4], [6], [7], [9]. Inspired by previous works, as a bandwidth, which is fundamental aspect of this approach, we use Scott's rule [11].

Dependency Indicators of Return Series

An essential part of portfolio theory is the issue of expressing the dependence structure of assets. Thus, to express a dependency structure return series r_i, we can use many approaches, e.g., correlation, copula functions, etc. [8]. The well-known approach used also in the portfolio risk measurement is variance–covariance. Mathematically, the covariance between the i-th and the j-th return is given by

$$\sigma_{i,j} = E[(r_i - \mu_i)(r_j - \mu_j)], \tag{4}$$

where E is the operator of the expected value and μ_i is the mean of the i-th returns.

Next, a commonly used linear dependency indicator is the Pearson coefficient of correlation, which is given by

$$\rho_{r_i,r_j}^{Pearson} = \frac{E[(r_i - \mu_i)(r_j - \mu_j)]}{\sigma_{r_i}\sigma_{r_j}}, \tag{5}$$

where μ_i is the mean value of r_i, and standard deviation $\sigma_{r_i} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_{i,t} - \mu_i)^2}$.

Several years ago, Ruttiens suggested a different perspective on risk as well as dependency measurement while incorporating the time (trend) factor, [10]. This concept is based on cumulative return $c_{i,t}$ calculated as $c_{i,t} = c_{i,t-1}\exp(r_{i,t})$ and linear trend line $e_{i,t}$ called "equally accrued return" leading to the identical final cumulative value, where $t = 1, 2, \ldots, T$. Additionally, e_i is simply computed as a linearly weighted return $e_{i,t} = c_{i,0} + \frac{t}{T}(c_{i,T} - c_{i,0})$, where $c_{i,0}$ represents the amount of initial investment. Recently, a modified version of the Ruttiens correlation measure was proposed by [4]. Following their theoretical concept, the modified trend depended correlation ρ_{r_i,r_j}^{TD} is formulated as:

$$\rho_{r_i,r_j}^{TD} = \frac{E\left[(c_i - e_i)\left(c_j - e_j\right)\right]}{\sigma_{(c_i - e_i)}\sigma_{\left(c_j - e_j\right)}}, \tag{6}$$

where the standard deviation of the spreads is calculated as $\sigma_{(c_i - e_i)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(c_{i,t} - e_{i,t})^2}$.

3 Empirical Analysis

This section first describes the data set used, then characterizes the empirical procedure, and finally presents the results with a discussion.

3.1 Data

For this study, we use active stocks (on date 4 November 2021) included in the S&P 500 index. In particular, the dataset consists of daily returns of stocks in the interval from 3 January 2005 to 4 November 2021. However, we split the time period into 3 sub-periods: Period A from 3 January 2005 to 31 December 2010, Period B from 3 January 2011 to 31 December 2017 , Period C from 1 January 2015 to 4 November 2021. The data is gathered from the Bloomberg Database. As a risk-free rate, the 3-month U.S. Treasury bill return is used in order to compute performance measures.

3.2 Empirical Process with Results

The portfolio selection framework considered in this analysis includes monthly recalibration of composition, i.e. each 21 trading day, while using a moving window of one year (252 trading days) historical observations of returns. Furthermore, short sales of assets are not allowed, and an upper limit of weight is not set. For simplification, we set the initial wealth $W_0 = 1$. Note that we exclude some of the stock series from the database due to the unavailability of data during selected interval. Once we do that, there remain 374 stocks in the data set.

The whole empirical procedure can be divided into following three steps:

Step 1. Determine the dependency matrices of returns considered variance–covariance (cov), Pearson correlation (corrP), and trend-dependent correlation (corrTD) based on 252 historical observations. Then, apply the PCA to particular dependency matrix and determine number of factors denoted as # (main components) using different strategies. In particular, we distinguish strategy with static number of components and dynamic number of components with required explanation power. Finally, approximate the future returns by RW regression model (2).

Step 2. Determine the optimal vector of asset weights x applying the Sharpe ratio (SR) maximization framework to approximated returns from the previous step. Again the rolling window of 252 days is considered. Also, the condition that the weight of one asset does not exceed 20% is imposed. The formulation of the optimization framework is as follows:

$$\begin{aligned} \max_{x} & \frac{E(x'r - r_f)}{(x'\Sigma x)^{\frac{1}{2}}} \\ & x'\varphi = 1 \\ & 0 \geq x_i \geq 0.2; \;\; i = 1, \ldots, z, \end{aligned} \tag{7}$$

where r_f is the risk-free return, Σ is the non-singular covariance matrix of asset returns, and φ is an z-column unit vector with all values equal to one.

Step 3. After determining the asset weight structure in the portfolio, calculate the ex-post final wealth (FW) and selected performance measures.[1] The whole process is repeated until the end of the time period is reached.

The results obtained by this analysis are presented in Figures 1, 2, and 4 for strategy with static number of factors and Table 1 for strategy with defined explanation power. Essentially, we present the effect on the final wealth of individual portfolios and their SR although we also calculated other indicators such as the mean, standard deviation, Rachev ratio and others as part of the analysis. Since the results leads the same conclusions, the rest of indicators were not presented here.

From the figures demonstrating the results of strategy with a static # during selected periods, we can observe that portfolio performance indicators are sensitive to the inclusion of a small number of factors. Furthermore, using more than 15 factors (approx. 4% of all factors) for approximation does not generate significant shifts in portfolio performance. Since this fact is valid in all periods, the portfolio results are then more robust. It also appears that in periods A and B there is no difference in these situations when the dependence matrix is changed. Only in period C it is advantageous to use corrTD. Additionally, the results of FW correspond to the performance measure SR, where the excess return is compared with the standard deviation [12].

Rather surprisingly, the best performance of portfolios is mainly achieved using less than 5 factors. In this situation, the corrP generates the best results of FW and SR mainly for periods A and C. In the next situations, where # is

[1] We do not include transaction costs in this analysis.

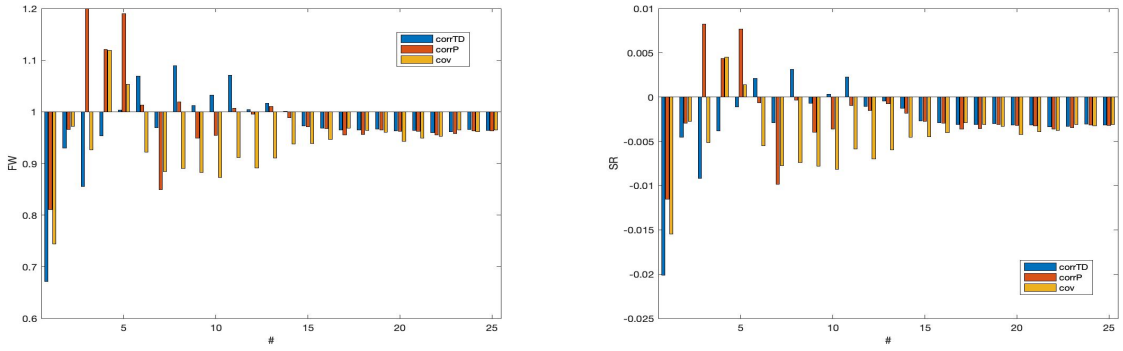


Figure 1: Selected portfolio statistics for different number of factors and dependency matrices to PCA in period A

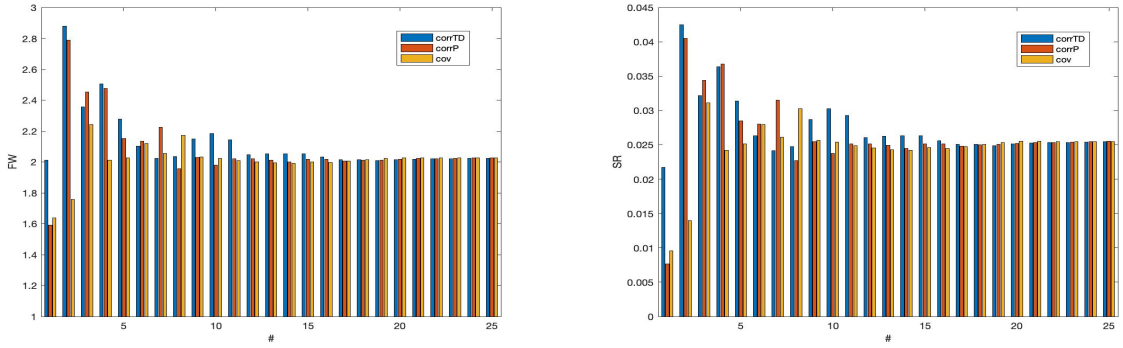


Figure 2: Selected portfolio statistics for different number of factors and dependency matrices to PCA in period B

in the interval approximately from 6 to 14, we can observe the benefits of employing corrTD to the PCA. This is especially visible in time periods A and C, where the market was volatile for a part of the period. Note that in these periods the FW of the portfolio fluctuates around 1, whereas in period B the value of the portfolio is approximately doubled.

The interesting point of this analysis is the explanation power of different dependency matrices used for the PCA. In all periods analysed, it can be observed that the set of factors obtained from the PCA to trend-dependent correlation matrix explains more overall portfolio variability. A brief discussion of trend-dependent correlation PCA properties was provided in paper [4], but without clearly presented results. Notice that the first factor with the highest explanatory power is obtained by including Pearson correlation to the PCA. However, jumps in the explanatory power of the second and third factors of the trend-dependent correlation result in a significant increase in the cumulative explanatory power. The differences between the remaining dependency matrices are not as pronounced.

In the next part, we use the second strategy considered in this analysis, where we consider a dynamic composition of factors explaining a certain portion of portfolio variability. For illustration, we select 5 intervals (70%, 75%, 80%, 85%, 90%), which have been discussed in the literature, [6]. The results are shown in Table 1.

From the results of this analysis in Table 1, it is evident that the differences between FW of constructed portfolios are not so pronounced, which is caused by the short-term intervals of specified investment periods. In period A affected by the global financial crisis, it is evident that some portfolios are below the initial investment threshold and SR shows negative values due to the excess return being negative. If we look in detail at the individual strategies, the results also show that using variance–covariance and Pearson correlation matrices to the PCA is less sensitive to the required explanation of portfolio variability mainly during the rising market period (period B). Furthermore, there are no differences in performance between the two measures, only in the average number of factors considered for PCA, which is higher for corrP.

As already evident from the previous analysis, corrTD uses significantly fewer factors to explain the required amount of total portfolio variability while at the same time, these portfolios outperform the others. Also, the overall portfolio performance is higher if a lower required level of explanation is preferred (70%).

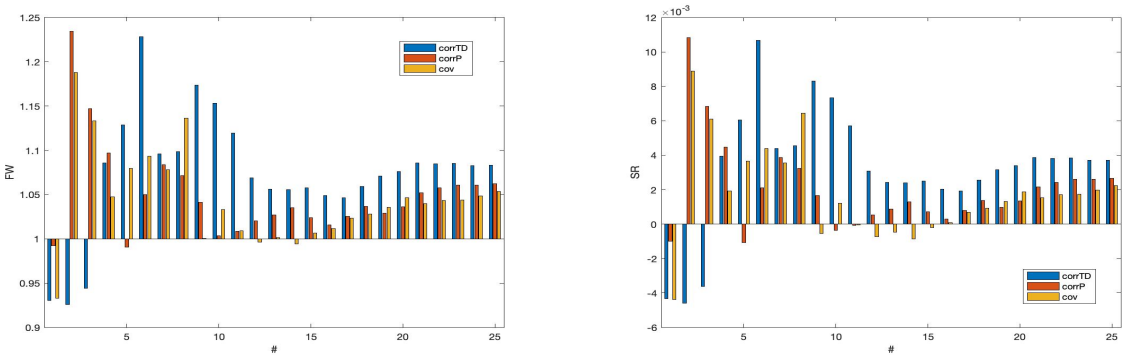


Figure 3: Selected portfolio statistics for different number of factors and dependency matrices to PCA in period C

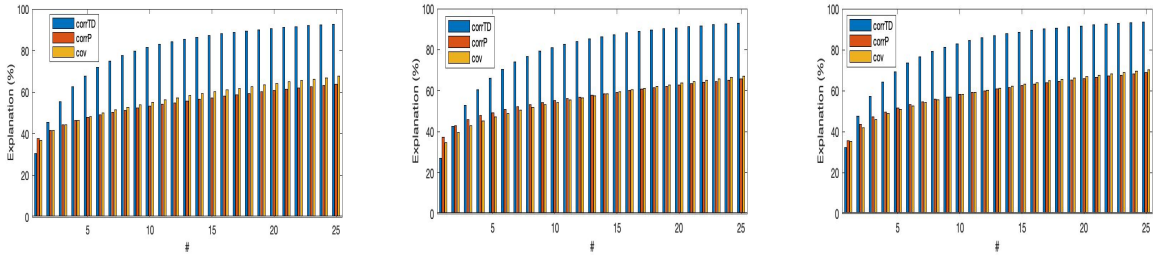


Figure 4: Cumulative mean explanation power of different types of dependency matrices used for PCA with respect to number of factors in periods A-C (ordered from left)

In general, according to this analysis, we can say that more robust results are when using more factors, but in this case investors can not achieve the best performance from the market components. However, the main empirical contribution is to introduce significantly useful properties of trend-dependent correlation for PCA.

In the further research, we could focus more on the datates used for the analysis from different markets. Moreover, it is possible to examine the trend–risk measures in more complex strategies presented by [2] or [13].

4 Conclusion

This paper provided analysis focused on the impact of factor number setting strategies used for return approximation in the portfolio selection process. Moreover, we also compared the effect of three different dependency matrices used for PCA in order to obtain main factors. In empirical analysis, we compared the portfolio performance statistics (FW and SR) while different strategies or numbers are considered. For these purposes, the SR maximization model was applied to the US stock dataset whilst three different periods were considered.

The results obtained from this analysis showed better explanation properties of the trend-dependent correlation matrix compared to the traditional ones. Furthermore, it was observed that the inclusion of more than 15 factors, while taking into account the original dataset, minimizes the differences in portfolio performance even though different dependence matrices are applied. The highest performance of the portfolios was achieved when approximately 5 factors were included. Note that the results were not very consistent for particular dependency matrix.

This paper mainly served as a sensitivity analysis to properly set the parameters of the return approximation problem in further portfolio selection research. In particular, we found that it may be more appropriate to build complex strategies on the properties of trend-dependent correlation PCA.

Acknowledgements

The author greatly acknowledged support through the Czech Science Foundation (GACR) under the project GA20-16764S and a SGS research project SP2023/019 of VSB-TU Ostrava.

	corrTD			corrP			cov		
expl	FW	SR	#	FW	SR	#	FW	SR	#
	Period A								
70%	1.001	-0.125	6.1	1.018	-0.040	36.6	0.903	-0.660	28.9
75%	0.985	-0.206	7.6	0.923	-0.547	47.8	0.973	-0.276	38
80%	1.103	0.373	9.6	0.937	-0.469	61.6	0.940	-0.453	49.9
85%	1.037	0.059	12.9	0.936	-0.474	78.9	0.936	-0.474	65.7
90%	0.996	-0.154	19.1	0.936	-0.475	102	0.936	-0.474	87.7
	Period B								
70%	2.410	3.481	6.4	2.002	2.463	33.2	1.955	2.293	29.6
75%	2.335	3.360	7.8	2.006	2.478	44.3	2.007	2.481	39.1
80%	2.016	2.461	9.9	2.007	2.482	58.2	2.007	2.482	51.1
85%	1.952	2.280	13.3	2.007	2.482	75.6	2.007	2.482	66.9
90%	2.019	2.516	19.2	2.007	2.482	99	2.007	2.482	89
	Period C								
70%	1.113	0.519	5.7	1.030	0.107	28.4	1.039	0.152	25.6
75%	1.101	0.471	6.9	1.063	0.276	38.2	1.097	0.448	34.1
80%	1.077	0.355	8.9	1.056	0.234	50.6	1.051	0.214	45.1
85%	1.045	0.186	11.9	1.079	0.350	67.1	1.051	0.208	60
90%	1.056	-0.154	17.4	1.066	0.287	89.7	1.067	0.290	81.4

Table 1: Selected portfolio statistics using dynamic strategy for different dependency matrices to PCA and selected time periods

References

[1] Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

[2] Bai, J. & Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 25(1), 52–60.

[3] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.

[4] Neděla, D., Ortobelli, S. & Tichý, T. (2023). Mean–variance vs trend–risk portfolio selection. *Review of Managerial Science*. https://doi.org/10.1007/s11846-023-00660-x.

[5] Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economic and Statistics*, 92(4), 1004–1016.

[6] Ortobelli, S., Kouaissah, N. & Tichý, T. (2017). On the impact of conditional expectation estimators in portfolio theory. *Computational Management Science*, 14(4), 535–557.

[7] Ortobelli, S., Kouaissah, N. & Tichý, T. (2019). On the use of conditional expectation in portfolio selection problems. *Annals of Operations Research*, 274(1), 501–530.

[8] Ortobelli, S. & Tichý, T. (2015). On the impact of semidefinite positive correlation measures in portfolio theory. *Annals of Operations Research*, 235(1), 625–652.

[9] Ruppert, D. & Wand, M. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3), 1346–1370.

[10] Ruttiens, A. (2013). Portfolio risk measures: the time's arrow matters. *Computational Economics*, 41(3), 407–424.

[11] Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley & Sons Inc.

[12] Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.

[13] Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association*, 113(523), 1341–1349.

Parameter Optimization of Trend Detection Algorithm Presented on Selected Stock Prices

Jakub Neugebauer [1]

Abstract. Trend analysis and prediction of stock prices have always been key parts of investment strategy decision-making. Many methods to support this analysis have been purposed from basic ones like technical analysis to more complex ones like machine learning methods, which have been lately on the rise in many econometric fields including time series prediction. The method of time series trend analysis used in this paper is based on community detection of a network created from sets of statistics describing trends at a certain time. The goal of this paper is to set up parameters of the already proposed trend detection algorithm in order to maximize the precision of trend detection. The function to maximize used in this paper is a total return from investments if a stock has always been bought at the uptrend beginning and sold (shorted) at the start of the downtrend. The behavior of the trend detection algorithm under its different parameters is shown on real data of selected stocks from S&P 500. Further inspection of algorithm predictive ability is shown on these data as well. The outcome of this paper is a method that finds optimal parameters of the trend detection algorithm and even further improves its precision.

Keywords: stock price prediction, trend detection, parameter optimization

JEL Classification: C44, B23
AMS Classification: 90C59, 62P05

1 Introduction

There are many factors, which can be looked at in order to make the best decision while investing in the best possible asset and achieving a return as highest as possible. One of the ways to do so is to predict the future price of a stock. Unfortunately, this is not an easy task as stock price development is a stochastic process and their price movements are determined by so many factors, that they are considered as random. There are many methods to predict the future price of a stock. Classical approaches are regression methods such as ARIMA models (see Box et al. [3]) and currently, machine learning methods (see Chen et al. [4]), such as neural networks (see Kim and Won [5]) are on the rise in this field.

Another way to decide which asset should the investor invest in is to analyze the trend. Many methods to do so have been developed (see Li and Liao [6]). The community detection-based trend detection algorithm proposed in Anghinoni et al.[1] has proven to be precise, although its parameters were set up without any further investigation. The goal of this paper is to propose new parameters and optimize them in order to even further improve the precision of the trend detection algorithm.

2 Community Detection Based Trend Detection Algortithm

The base algorithm that is being looked at in this paper is the community detection-based trend detection algorithm proposed in Anghinoni et al.[1] This algorithm is originally divided into 4 steps but another step placed at the beginning is added in this paper and it is simple exponential smoothing of an original time series. The second step is the extraction of series features at each time. This step is a mathematical analogy to technical analysis, as it is aggregating several descriptive statistics into a vector describing the behavior of a time series at a specific time. This is the base of the trend detection algorithm, as its goal is to separate similar features into communities with similar trends. The third step of this algorithm is network generation, the fourth step is community detection on this network and finally the last step is trend detection on found communities. All these steps are described in the following section.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, nám. Winstona Churchilla 1938/4, 120 00 Praha 3-Žižkov, jakub.neugebauer@vse.cz

2.1 Exponential Smoothing

As was already written, the first step of the trend detection algorithm is the basic exponential smoothing of the original time series. This step was added to an original algorithm to get rid of problems with high volatility. Exponentially smoothed series x_t^s is derived from original time series x_t with smoothing parameter α. This parameter will be further used as an input in the parameter optimization

2.2 Feature Extraction

The most important part of the algorithm is deriving features from smoothed time series x_t^s. Basically, any statistic can be used as a feature. In this paper, the same features are used as in the original one. Three statistics are taken into consideration. The first one is noise, the second characteristic is gradient and the last statistic is the relative high-low position. For all these statistics two time frames, short-term (ST) and long-term (LT) are taken into consideration. Let q be a length of a time frame. Noise n_t^q at time t is calculated as follows:

$$n_t^q = \frac{x_t^s - MAq_t}{MAq_t} \tag{1}$$

Where MA_q is the moving average and q denotes the number of last observations taken into its calculation. In this case, $q = ST$ for the first feature $f_t^1 = n_t^{ST}$ and $q = LT$ for the second feature $f_t^2 = n_t^{LT}$. With this setup, chosen short-term and long-term time frames can be included in the parameter optimization of this algorithm.

Gradient g_t^q at time t is calculated as follows:

$$g_t^q = \frac{MAq_t - MAq_{t-1}}{MAq_t} \tag{2}$$

The third feature is calculated correspondingly as $f_t^3 = g_t^{ST}$ and the fourth feature as $f_t^4 = g_t^{LT}$.

The last characteristic taken into consideration is the relative high-low position rht_t^q and it is derived by this formula:

$$rht_t^q = \frac{x_t^s - \min\limits_i(x_i^s)}{\max\limits_i(x_i^s) - \min\limits_i(x_i^s)}, \quad i \in \mathbb{Z}, i \geq t - q \wedge i \leq t \tag{3}$$

This equation is a normalization to how close is the value of smoothed series x_t^s at time t to the maximum value in the last period of length q. Last two features are calculated again as $f_t^5 = rht_t^{ST}$ and $f_t^6 = rht_t^{LT}$.

All these features are in the end normalized by z-score and discretized, so they are in the interval $[0, 1]$. This interval is separated into n_b bins. The value of each feature is then assigned to the closest bin. For example, if the number of bins would be 3, each feature value would be in $\{0, 0.5, 1\}$. By this separation, each feature is categorical and points in time can be grouped together based on similar features. Normalized, discretized, and categorized features are then put into 6-dimensional vectors $F_t = (f_t^1, f_t^2, f_t^3, f_t^4, f_t^5, f_t^6)^T, \quad \forall t \in 1 \cdots T$, where $T = T' - LT$ is the number of created factors vectors, as some of the observation of original series of length T' were lost due to smoothing and usage of moving averages in feature extraction.

2.3 Network Generation

The third step of the trend detection algorithm is another algorithm itself. It is a network generation algorithm based on extracted features from the last step. The first part of this algorithm is assigning node labels to original nodes based on feature values F_t. In the second step, a network is created, which has labels as nodes, and the arch is between node i and node j if the original nodes are corresponding to two following timestamps. Let original nodes N_t feature be F_t, label node of N_t is $L(N_t)$ and adjacency matrix of the desired network is M_{xy}. The algorithm of network generation is described in algorithm 1.

In this way, weighted graph G with adjacency matrix M_{xy} is created, as the same two labels may continue each self multiple times in time. Weights are a number of repetitions of these occurrences.

2.4 Community Detection

In the last step, weighted graph G was created and the task in the fourth step of a trend detection algorithm is to separate nodes of this graph into communities. A group of nodes in one community is connected at a higher rate

Algorithm 1 Network generation

```
for t ∈ 1, ···, T do
    N_t = F_t
end for
label ← 1
for t ∈ 1, ···, T do
    if L(N_t) = ∅ then
        L(N_t) ← label
        for j = t + 1, ···, T do
            if N_t = N_j then
                L(N_j) ← label
            end if
        end for
    end if
    label = label + 1
end for
M ← 0_{label−1}
for t ∈ 2, ···, T do
    if L(N_{t−1}) ≠ L(N_t) then
        x = L(N_{t−1}),    y = L(N_t)
        M_xy = M_xy + 1
    end if
end for
```

and the weight of edges between the nodes have higher weights than between nodes from different communities if such edges exist. The community detection algorithm used in this paper is a widely used Louvain algorithm created by Blondel et al. [2]. To the parameter optimization, the resolution parameter res is added from this algorithm. The problem with separating nodes representing feature vectors is, that it is not straightforwardly visible, what is the trend of each community. Communities are created on labels of nodes, and therefore trend $trend_t$ at time t of original series x_t is equal to the trend of a community of label of node N_t i.e. $trend(x_t) = trend(com(L(N_t)))$.

2.5 Trend Classification

To determine the trend of each community found in the last step, a simulation of a walk through the timeline has to be done. Percentage change in original time series x_t is saved, if two consecutive nodes N_{t-1}, N_t are in the same community ($com(L(Nt-1)) = com(L(N_t))$). Then the average percentage change in each community is calculated and the community trend is determined and therefore trend at each time. On top of the original paper, two more parameters, the upper bound parameter UB and lower bound parameter LB, are added to the trend classification algorithm and a neutral trend (NE) is considered. Let C be the set of all communities found in the last step, Ω_{com} be the set of all percentage changes in the community com, Δ_{com} be the average percentage change in the community com and $trend(com)$ is the trend of community com. This algorithm is described in 2.

With this step, the trend detection algorithm ends and three different trends, up trend ('UP'), neutral trend ('NE'), and downward ('DW') are found. From this step, parameters UB and LB are added to parameter optimization, so the optimal differentiation between a possibly profitable trend ('UP' or 'DW' if difference between buying and selling price is higher than the spread) and a non-profitable trend ('NE') is found.

3 Parameter Optimization

To determine total return, a simulation of trading strategy is proposed, such that stock is bought at the beginning of the 'UP' trend and shorted at the beginning of the 'DW' trend with an initial budget equal to 1. If the trend is neutral, the position is unchanged. Spread is also taken into consideration in calculation, to avoid too frequent buying/selling operations. The stopping criteria is, that budget cannot get under 0, which means, that once all money is lost, the final budget is equal to 0. Let U be the set of ordered pairs, where the first element of this pair is an index of a time, where is stock bought with a supposed trading strategy and the second element is an index of a time, where this stock was sold. Accordingly, D is a set of ordered pairs, where the first element is an index of where the stock was shorted by trading strategy and the second element is a time when the short position was closed. The final budget B determined by this strategy is then:

Algorithm 2 Trend classification

for $com \in C$ **do**
  $\Omega_{com} \leftarrow \emptyset$
  $\Delta_{com} \leftarrow 0$
  $trend(com)$ = 'NE'
  for $t \in 2, \cdots, T$ **do**
    if $com(L(N_{t-1})) = com(L(N_t))$ **then**
      $\Omega_{com} \leftarrow \Omega_{com} \cup \frac{x_t}{x_{t-1}}$
    end if
  end for
  $\Delta_{com} = mean(\Omega_{com})$
  if $\Delta_{com} \geq UB$ **then**
    $trend(com)$ = 'UP'
  else if $\Delta_{com} \leq LB$ **then**
    $trend(com)$ = 'DW'
  end if
end for
for $t \in 1, \cdots, T$ **do**
  $trend(x_t) = trend(com(L(N_t)))$
end for

$$B = \prod_{(i,j)\in U} \left(\max \left(0, 1 + \frac{x_j(1-0.5spr) - x_i(1+0.5spr)}{x_i(1+0.5spr)} \right) \right) \prod_{(i,j)\in D} \left(\max \left(01 + \frac{x_i(1-0.5spr) - x_j(1+0.5spr)}{x_j(1+0.5spr)} \right) \right) \tag{4}$$

where spr denotes spread. In this way, it is possible to optimize the parameters of the trend detection algorithm in order to maximize return ret from function 4. In this paper, parameters are optimized for all chosen time series at once. Aggregation of returns from single stocks must be done in order to do so. Average daily return is derived by the geometric mean of returns and total ret is calculated as an average of average daily returns of all considered stocks.

$$MAX \quad ret = \frac{\sum_{i=j}^{n} (\sqrt[T_i-1]{B_i} - 1)}{n} \tag{5}$$

Where return ret_i of stock i is calculated based on trends derived from the trend detection algorithm, n is the number of the stocks which are used for parameter optimization, and T_i is the length of time series i after feature extraction. All the parameters are also bounded. Parameter $\alpha \in [0.05, 0.95]$, $ST \in \mathbb{Z}$, $ST \in [5, 100]$, $LT \in \mathbb{Z}$, $LT \in [10, 200]$, $n_b \in \mathbb{Z}, n_b \in [2, 10]$, $res \in \mathbb{Z}, res \in [2, 15]$, $UB \in [0, 0.0001]$ and $LB \in [-0.0001, 0]$. These bounds were set up in order to omit overfitting on the training data set. Spread was set to 0.4 % to represent reality.

3.1 Results

For the showcase of parameter optimization of the trend detection algorithm, 50 stocks from S&P 500 index were randomly selected. Time series consists of daily closing prices of these stocks from the beginning of the year 2010 till the end of April 2023 and are downloaded from Yahoo finance [7]. This range provides enough data for parameter optimization. Data are separated into a training data set $train$, which ends with the year 2020, and a test data set $test$ which consists of the rest of observed range.

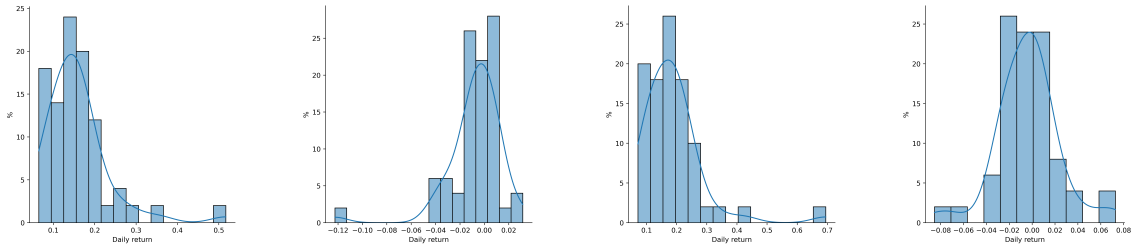
Prediction on the testing data set is done in a way, that feature vectors at each time are extracted and for features vectors, that already exist in the training data set, the same trend is assigned. For all other feature vectors, the neural network model with 8 hidden layers and the last activation function *softmax* is trained on the training data set and used to predict the trend.

The differential evolution method was used to optimize parameters. Found optimal parameters, as well as returns on training data set are written in table 1. For comparison with the original paper, parameters, and returns are compared with parameters proposed by *Anghinoni et al.* As in the original paper parameters α and res were not included, their values were set the same as in found best values. Parameters UB and LB were also not included by *Anghinoni*, but by setting them to 0, the original method is gained.

	Optimal	*Anghinoni et al.*
α	0.934	0.934
ST	6	25
LT	27	125
n_b	8	9
res	15	15
UB (e^{-5})	8.824	0
LB (e^{-5})	-1.590	0
ret_{train} (%)	0.187	0.159
ret_{test} (%)	-0.004	-0.008

Table 1 Sample Table

In the optimal parameters setup, only 8 bins, instead of the originally proposed 9 bins are used, the short-term period is lowered from 25 to 6 days, and the long-term period is lowered from 125 to 27. This suggests, that stock prices are more volatile than the time series used in the original paper. The average daily return is improved on the training data set from 0.159 % to 0.187 % as well as on the test data set, where the average daily return is improved twice from -0.008 % to -0.004 %. Distributions of returns on the training set are similar, as it is shown in figure 1. On the testing set, on the other hand, the distribution of returns on parameters by *Anghinoni et al.* is significantly left-skewed while on *Optimal* parameters, the distribution of returns is normal and suggests, that this parameter setup is more robust against extreme loses.



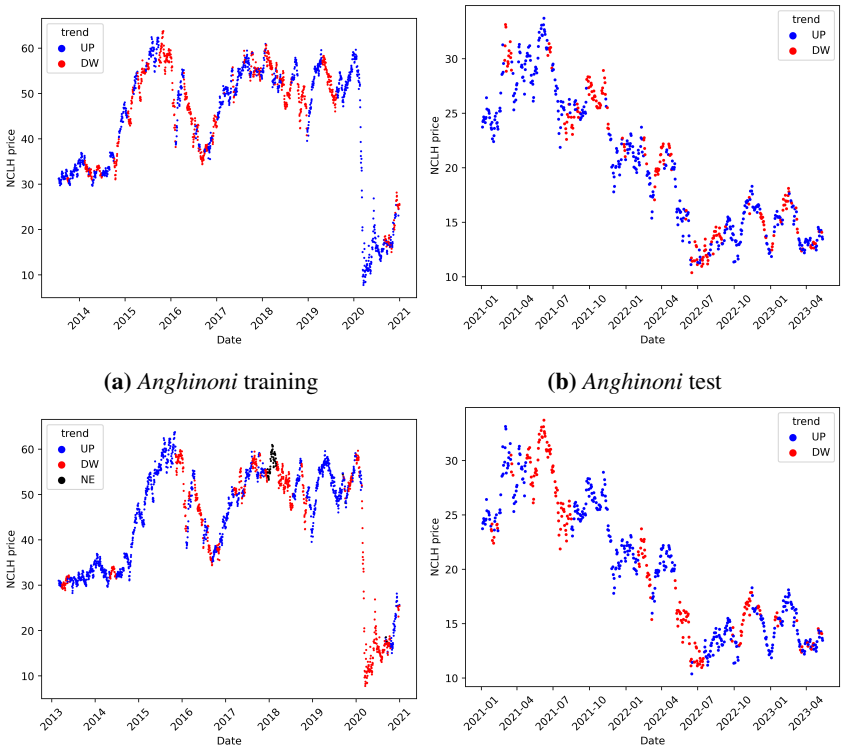
(a) *Anghinoni* training **(b)** *Anghinoni* test **(c)** *Optimal* training **(d)** *Optimal* test

Figure 1 Distribution of average daily returns

To show the efficiency of the trend detection algorithm, prices of Norwegian Cruise Line Holdings (NCLH) stock colored by detected trend is shown in figure 2. This stock was selected, because it is the one with the highest return on the test set using *Optimal* parameters. It seems that parameters proposed by *Anghinoni et al.* detect trends in shorter periods and more frequently on NCHL stock prices. On the other hand, it has not detected a huge downfall at the beginning of the year 2020 caused by the COVID-19 pandemic as a downtrend which is detected by *Optimal* parameter setup. Trend prediction seems also more stable with the usage of *Optimal* parameter setup. In both cases, there are occurrences of single points, which may disrupt the proposed trading strategy. The inefficiency of using trading strategy to get positive income confirms negative average returns on training data set in table 1. To use a trend detection algorithm as a tool for opportunity trading, some more sophisticated strategy should be applied. A simple strategy could be to hold a position for multiple days after the trend change and wait if these days are detected with the same trend.

4 Conclusion

The goal of this paper was to find the optimal parameters setup of the trend detection algorithm to improve its precision. To do so, an original algorithm was updated with new parameters, concretely smoothing parameter α, resolution parameter of Louvain algorithm res and lower (LB) and upper bound (UB) parameters, to differentiate also the neutral trend. All the steps of the trend detection algorithm were described. A model for parameter optimization was proposed. Its goal was to maximize the average daily returns of stocks. 50 random stocks from S&P 500 index were chosen to show the capability of the trend detection algorithm and the improvement of its precision after optimizing parameters. The average daily return has improved from 0.159 % to 0.187 % on the training data set and it has risen from -0.004 % to -0.008 % on the test set. After observing the behavior of trend



(a) *Anghinoni* training

(b) *Anghinoni* test

(c) *Optimal* training

(d) *Optimal* test

Figure 2 Detected trends of NCLH prices

prediction, it has been concluded, that a simple strategy to buy/short stock at trend change points is not sufficient to gain income and a more sophisticated strategy, like holding the position, until multiple days with opposite trends occur, should be used. Nevertheless, the precision of the trend detection algorithm has been increased by the proposed parameter optimization.

Acknowledgements

The research project was supported by Grant No. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] Anghinoni, L., Liang, Z., Donghong, J. & Heng, P. (2019). Time Series Trend Detection and Forecasting Using Complex Network Topology Analysis. *Neural Networks*, 117, p. 295–306.

[2] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, p. P10008

[3] Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons Inc.

[4] Chen, W., Zhang, H., Mehlawat, M. K. & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, 100, 106943.

[5] Kim, H. Y. & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 103, p. 25-37.

[6] Li, W. & Liao, J. (2017). A comparative study on trend forecasting approach for stock price time series. *In 2017 11th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, 74–78

[7] Yahoo Finance (2023). Finance Home. [online] Available at: https://finance.yahoo.com/, [cited 2023-05-13]

Models of Military Expenditures

Jakub Odehnal[1], Lenka Brizgalová[2], Jiří Neubauer[3], Lucie Svobodová[4]

Abstract. The aim of the article is to present the possible use of the Autoregressive Distributed Lag model to identify military expenditures determinants and to verify the model's ability to make short-term predictions of military expenditures of selected Baltic countries. In order to analyse economic environment as determinant of military expenditures, the following variables were used: budget balance as percentage of gross domestic product, foreign debt as percentage of gross domestic product, inflation, the state of the economy measured by gross domestic product. The results do not reveal uniformity of economic determinants among the Baltic countries, but confirm the positive effect of gross domestic product on military expenditure in case of Latvia and Estonia. The positive effect between military spending and inflation is confirmed for Lithuania. In case of Lithuania, Latvia and Estonia, the expected positive link is found between military expenditures and the levels of fiscal deficit or external debt. Generally, we can conclude that the economic environment is one of the main factors influencing military expenditures; however, at present, military expenditures are mainly driven by the change in perception of security threats in Europe.

Keywords: ARDL model, Economic Determinants, Military Expenditures

JEL Classification: C33, E69
AMS Classification: 62J05

1 Economic Determinants of Military Expenditures

Over the past year, a significant increase in military spending can be observed as a part of government spending in most Alliance countries, partially as a result of the changed security situation in Europe, and also as a manifestation of the conclusions of the 2014 Alliance Summit in Wales. In general, in addition to changes in the security environment, changes in the economic environment, manifested primarily through changes in a country's economic development as measured by the size and change in gross domestic product (GDP) and fiscal indicators describing the size and change in fiscal deficit or the level of external debt can also be considered as factors affecting the amounts allocated to military spending. The authors of the paper analyse the determinants of military expenditures of Lithuania, Latvia and Estonia as the NATO member countries with the highest growth rates of military expenditures in the analysed period. Moreover, these are examples of countries that are willing to fulfil the political commitment to pay 2 % of GDP on defence. The analysis of the so-called determinants of military expenditures is the subject of a number of empirical studies [12, 11, 18, 14, 17, 3, 4, 1, 16, 19], where authors usually describe the connections between the amount allocated to a country's military expenditures and selected economic variables as their determinants.

Using Egypt as an example, Abdelfattah et al. [1] analyse the link between military spending and a country's economic strength as measured by gross domestic product, the levels of military spending of neighbouring countries (Israel, Jordan, Syria), the size of the population, the quality of democracy, and the degree of economic openness of the country as measured by the share of net exports in the country's GDP. The findings show that both economic and strategic factors play a role in determining the level of military spending, with the level of military spending from the previous year in particular having a clear positive effect on military expenditures. Given the influence of Israel's military spending, the strategic effect is mostly positive; however, in general, the military spending of Jordan and Syria has no effect on Egyptian military expenditures. The conclusions of a study carried out by Nikolaidou [11] confirm a positive relationship between the size of gross domestic product and the size of military spending for seven of the countries studied. Similar conclusions about the influence of economic performance on military expenditures are shown in Solarin's work [18], which is based on an analysis of 82 countries carried out over the period of 1989-2012. In their paper, Smyth and Kumar Narayan [17] analyse the effect of

[1] University of Defence, Brno, Department of Resource Management, Kounicova 65, Brno, jakub.odehnal@unob.cz
[2] University of Defence, Brno, Department of Resource Management, Kounicova 65, Brno, lenka.brizgalova@unob.cz
[3] University of Defence, Brno, Department of Quantitative Methods, Kounicova 65, Brno, jiri.neubauer@unob.cz
[4] University of Defence, Brno, Department of Resource Management, Kounicova 65, Brno, lucie.svobodova@unob.cz

military spending on foreign debt for six countries—Oman, Syria, Yemen, Bahrain, Iran, and Jordan between 1988 and 2002. The authors conclude that a 1% increase in military spending leads to an increase of the country's foreign debt by almost 1.6% from the long-term perspective, and by 0.2% from the short-term perspective. Dunne, Perlo-Freeman, and Soydan [4] consistently show a positive link between the amount allocated to military expenditures and the level of a country's foreign debt. The authors carried out an analysis based on Argentinian, Chilean, and Brazilian economies, and applied the Autoregressive Distributed Lag (ARDL) model. The results of their analysis correspond to those obtained by Smyth, Narayan [17], and show a positive link between the studied variables. Author Yalta [19] emphasizes the inclusion of security determinants of military expenditure into an econometric model analysing selected eight countries in the Gulf Region (Saudi Arabia, Iran, Iraq, Kuwait, Bahrain, Qatar, UAE and Oman). Findings based on data for the period between 1980 and 2016 indicate that military expenditures are influenced jointly by economic and strategic determinants.

In the following part of the paper, the authors will analyse the influence of selected determinants of military expenditures based on the variables used by the above-mentioned authors [12, 11, 18, 14, 17, 4, 1, 16, 19]. The model for estimating the demand for military spending will be constructed using the following economic variables—GDP, inflation, external debt, and fiscal deficit. A regression model with lagged value of the explained variable (ARDL model) was used in order to analyse the association between the selected determinants of military expenditures and their size.

2 The Evolution of Military Expenditures

Military spending is a measure of economic resources allocated to the army. It is expressed in the country's current prices, in US dollars in constant prices, or as a share of the country's gross domestic product. The latter is the most accurate representation of the burden the military sector poses on the government budget. In the second chapter of this paper, the history of military spending as percentage of GDP in selected countries is analysed over the period between 1993 and 2021. The purpose is to find out whether the countries surveyed were spending the required 2% of GDP on their defence. The countries surveyed are Lithuania, Latvia, and Estonia, which joined NATO on 29 March 2004.

During the aforementioned period, Estonia spent the most on military expenditures, which ranged from 0.76% to 2.31% of GDP. Lithuania spent the least, with military spending as a percentage of GDP ranging from 0.45% to 2.12%. Lithuania saw a decline in military spending in 1994, 1995, 1998, and 1999. After the country became a member of NATO, there were annual increases until 2009, when military spending declined due to the global financial crisis. Latvia reports a decline in military spending between 1995 and 1997. From 1997 to 2008, their military spending increased gradually. Latvia's resource envelope ranged from 0.58% to 2.30% of GDP. In 2008, there was a decline which persisted until 2013. Between 2009 and 2014, the defence budgets of the Baltic states were impacted by the global financial crisis. This was most evident in Latvia and Lithuania. In Latvia, cuts were made in the command structure; for example, up to 40% of positions were cancelled, and salaries were reduced by 20% [9]. The most important decision taken by the Latvian government in response to Russia's military aggression against Ukraine was to increase military spending to 2% of GDP in 2018. Nevertheless, Latvia accentuates so-called comprehensive defence, which means that the whole society and other government institutions are tasked with defending the country. It also means that a significant part of the budget of other ministries is actually spent on defence, but these funds are not figured in the target 2% of GDP or the Ministry of Defence (MoD) budget. [2]

In case of Estonia, the slight reduction in defence spending during the global financial crisis affected mainly the navy, which lost a significant part of its skills. [6] As for the army staff, there has been a reduction in senior positions; however, as they agreed in 2015, all political parties are still committed to achieving and progressively exceeding 2% of GDP in defence spending. Between 2019 and 2020, purchases of military equipment, infrastructure costs and investment costs accounted for more than one third of Estonia's defence budget. [5]

The Lithuanian armed forces had to deal with a reduction in their budget during the global financial crisis, which became apparent between 2009 and 2013. Russia's aggression against Ukraine in 2014 made Lithuania increase its military budget. In January 2017, the Lithuanian Parliament adopted a new National Security Strategy, and committed to allocating 2% of GDP in 2018 and to gradual increases in military spending. In September 2018, the majority of Lithuanian parliamentary parties signed an agreement regarding guidelines for Lithuanian defence policy, committing to allocate at least 2.5% of GDP to military spending in 2030. [8]. The Baltic states confirmed their plans to increase military expenditures to 3% of GDP in the upcoming years. Estonia, for example, has already confirmed plans to spend EUR 1 billion on defence over the next four years, roughly 2.9% to 3% of GDP. In 2022, Lithuania was supposed to spend about USD 1.3 billion on defence, which was more than 2% of GDP, but due to

the fact that Russia threatened to occupy Lithuanian territory in order to gain access to the Baltic Sea, the Lithuanian government will propose the parliament increase defence spending to 2.5% of GDP in the spring. The Latvian government also supports gradual increases in the defence budget over the next three years—up to 2.5% of GDP by 2025. The budget is projected to be at least 2.4% of GDP in 2024, and at least 2.5% of GDP in 2025 and beyond. [9]

At the Wales Summit in September 2024, government spokespersons of NATO (North Atlantic Treaty Organization) member countries pledged to gradually increase military budgets to reach two percent of gross domestic product by 2024 at the latest. Estonia has been fulfilling this commitment since 2015. Lithuania and Latvia only increased their budgets to reach two percent in 2018, and as of this year, all three Baltic countries are meeting the alliance's commitment to allocate 2% of GDP to defence.

3 Data and Methods

The article uses data on the development of military spending as a share of gross domestic product and its size in millions of US dollars in constant 2020 prices [15] in Lithuania, Latvia and Estonia. The economic determinants used were the gross domestic product in billions of US dollars at constant 2020 prices [15], gross domestic product growth in percentage points [15], inflation in percentage points [15], fiscal deficit as percentage of GDP [15], and external debt as percentage of GDP [7]. This paper presents an analysis of military expenditures of only three selected, newly admitted NATO member countries between 1993 and 2021, and also interprets the obtained results. These countries were selected based on their highly dynamic changes in military spending. Data available from the Stockholm International Peace Research Institute (SIPRI) international database and Organisation for Economic Co-operation and Development (OECD) were used to analyse the datasets.

The impact of the determinants of military spending on the amount of military spending in each country will be analysed using a linear regression model with a lagged explanatory variable of the form of

$$MILEX_t = \beta_1 + \beta_2 MILEX_{t-1} + \beta_3 GDP_t + \beta_4 INFL_t + \beta_5 DEF_t + \beta_6 DEBT_t + \varepsilon_t, \qquad (1)$$

where $MILEX_t$ represents military expenditures, $MILEX_{t-1}$ is lagged value of military expenditures, GDP_t represents GDP, $INFL_t$ is inflation, DEF_t is fiscal deficit, and $DEBT_t$ represents external debt.

4 Results

Table 1 contains the results of military expenditure demand models for Lithuania, Latvia, and Estonia. For each of the countries analysed, a model containing mainly statistically significant regressors is presented. The standard errors of the estimates are given in parentheses.

Variable	Lithuania	Latvia	Estonia
Intercept	-0.053 (0.054)	-0.045 (0.056)	-0.194*** (0.050)
Lagged value of military expenditure ($MILEX_{t-1}$)	1.010*** (0.054)	0.892*** (0.117)	0.492*** (0.129)
GDP (GDP_t)		0.006* (0.004)	0.017*** (0.004)
Inflation ($INFL_t$)	0.007** (0.003)		
Fiscal deficit (DEF_t)	0.013*** (0.004)	0.017*** (0.005)	
External debt ($DEBT_t$)	0.003** (0.001)		0.005** (0.002)
R^2	0.963	0.937	0.984
Adjusted R^2	0.956	0.929	0.982

Note: *p <0.1, ** p <0.05, ***p <0.01.

Table 1 Estimated models

The long-term relationship for Lithuania can be described by the following equation:

$$MILEX_t = -0.053 + 1.010MILEX_{t-1} + 0.007INFL_t + 0.013DEF_t + 0.003DEBT_t \quad (2)$$

The long-term relationship for Latvia can be described by the following equation:

$$\widehat{MILEX}_t = -0.045 + 0.892MILEX_{t-1} + 0.006GDP_t + 0.017DEF_t \quad (3)$$

The long-term relationship for Estonia can be described by the following equation:

$$\widehat{MILEX}_t = -0.194 + 0.492MILEX_{t-1} + 0.017GDP_t + 0.005DEBT_t \quad (4)$$

Based on the residual analysis of the above-mentioned models, the residuals can be considered normally distributed, uncorrelated and stationary (Shapiro-Wilk test [p-values: Lithuania 0.18673, Latvia 0.54355, Estonia 0.15828], Jarque-Bera [p-values: 0.46868, 0.60690, 0.46540] normality tests were used; autocorrelation was tested using Ljung-Box test [p-values: 0.39225, 0.54907, 0.17523], stationarity by ADF test [p-values: <0.01, 0.021831, < 0.01] and KPSS test [p-values: >0.10, >0.10, >0.10]).

The estimated parameters of the linear regression model described in Table 1 indicate that in case of Lithuania, military spending is affected by the amount allocated to military spending in the previous year, the level of inflation, fiscal deficit and external debt. As for Latvia, the military expenditures are influenced by the variable amount of military expenditures in the previous year, gross domestic product, and fiscal deficit. As far as Estonia is concerned, the military spending of the previous year, gross domestic product and the external debt proved to be the deciding factors.

Based on the estimated model (2) and the known values of the explanatory variables for the year of 2021, a forecast of the value of Lithuania's military expenditures for the year 2021 was created, $\widehat{MILEX}_{2021} = 1.309$ mld. USD, at the 95% confidence level [1.155, 1.463]. The actual military spending of 2021, which was USD 1.160 billion, is lower than the value estimated by the model, but still within the 95% confidence level.

Model (3) was used to forecast Latvia's military expenditures for 2021, $\widehat{MILEX}_{2021} = 0.686$ mld. USD, at the 95% confidence level [0.557, 0.815]. The actual military spending of 2021, which was USD 0.774 billion, is higher than the value estimated by the model; however, it is still within the 95% confidence level.

Based on the estimated model (4) and the known values of the explanatory variables for the year of 2021, a forecast of the value of Estonia's military expenditures for 2021 was created, $\widehat{MILEX}_t = 0.773$ mld. USD, at the 95% confidence level [0.710, 0.837]. The actual military spending of 2021, which amounted to USD 0.705 billion, is lower than the value estimated by the model, which means it is just below the 95% confidence level.

The amounts of military expenditures of the analysed countries, together with the values obtained using the estimated models (2), (3) and (4), and the forecasts are shown in Figure 1.

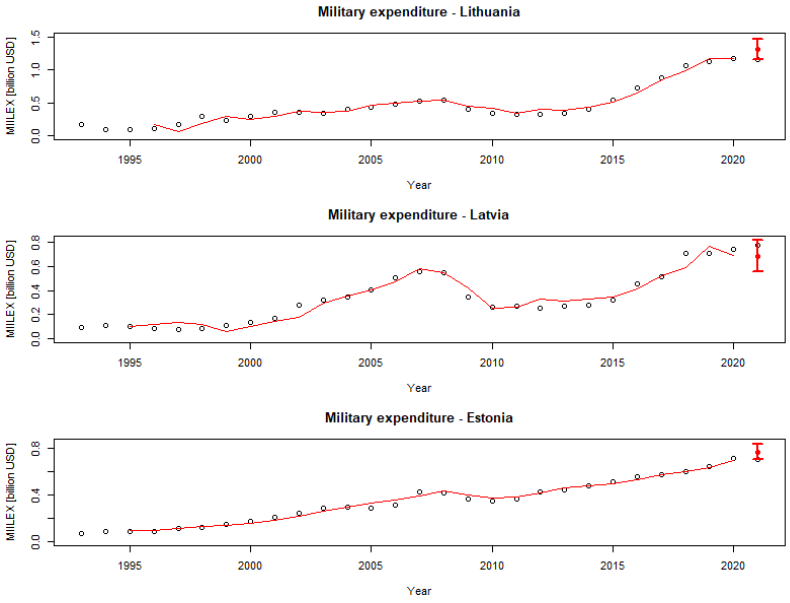


Figure 1 Military expenditure in Lithuania, Latvia, and Estonia (actual values, fitted and forecasted values)

5 Conclusions

Lithuania, Latvia, and Estonia are examples of the few NATO member states whose economies are experiencing rising military spending and, in addition, are fulfilling the political commitment to allocate 2% of GDP to defence. The result of the analysis showed that during the examined period (1993-2021), all three above-mentioned countries experienced extreme increases in military spending as a result of certain economic and security factors. Nevertheless, the global financial crisis starting in 2008 had a negative impact on this trend not only in the Baltic states, but also in the majority of NATO member countries. After annexation of Crimea by the Russian Federation in 2014, the security situation of the Baltic countries changed, especially because in terms of security, they are the most vulnerable NATO members.

The results of the regression analysis show that the expected positive association between GDP and military expenditures, as in the case of the analysis carried out by Nikolaidou [11], is both statistically significant and strong in the case of Latvia and Estonia, with both Latvia and Estonia experiencing long-term increases in military expenditures over the period analysed, thus confirming this association. The expected positive link between inflation and the amount of military spending was confirmed in case of Lithuania, where military spending can be expected to increase in a period of rising inflation in order to maintain the real purchasing power of the military. The same positive link between inflation and the amount allocated to military spending was confirmed in a study by Smyth and Narayan [17]. In case of the size of the fiscal deficit (surplus) and the amount of military expenditures, the expected (positive) link in case of Lithuania and Latvia was confirmed, which leads to the conclusion that a possible increase in the surplus acts as a factor increasing the size of military spending. When examining the relationship between external debt and military spending, a strong positive relationship can be observed in case of Lithuania and Estonia, confirming the theoretical conclusions of Smyth, Narayan [17], and Dunne, Perlo-Freeman, and Soydan [4] regarding the possible influence of military spending on external debt.

The nature of the datasets clearly shows a negative trend in military spending during the global financial crisis and the global COVID-19 pandemic, and a subsequent significant increase in military spending related to changes in the security situation towards the end of the analysed period. In terms of economic determinants, economic policy aimed at preparation for the adoption of the single European currency seems to play an important role. Unlike in most of the new NATO member countries, measures aimed at reducing the fiscal deficit did not limit the military spending of the Baltic states, but led to implementation of their responsible defence policy of increasing military spending and fulfilling their declared political commitment to NATO.

Using the ARDL method, it was found out that in Lithuania, the amount of military spending was influenced by inflation, fiscal deficit, and external debt. As for Latvia, military expenditures were influenced by GDP and fiscal deficit. In case of Estonia, GDP and external debt played an important role. The results of the regression analysis, which proved to be a suitable tool, and subsequent prediction confirmed an extreme increase in military spending for all three countries studied. For Lithuania, the predicted value for 2021 was slightly higher than the actual value, but within the confidence interval. Latvia's predicted value for 2021 was lower than the actual value, but it also fell within the confidence interval, and proved to be the most inaccurate of the three models examined. The difference between the predicted and the actual values was only USD 0.08 billion. For Estonia, the predicted value for 2021 was the only one higher than the actual value, and in this case it was not just below the confidence level. The values for Lithuania and Latvia speak in favour of the constructed model, the values for Estonia testify against the constructed model. Considering the fact that the models for all three Baltic countries are based only on economic variables, and the security environment does not play a role, the results speak in favour of the constructed models in two out of three cases. Possible research can continue with the analysis of strategic (namely the military expenditures of both the rival and the ally countries) and institutional factors influencing the size of military expenditures, which can extend the developed model and contribute to the improvement of short-term predictions.

Acknowledgements

The paper was supported by institutional funding aimed at the development of the research organization (Long-Term Organisation Development Goals at the FML).

References

[1] Abdelfattah, Y. M., Abu-qarn, A. S., J. Dunne, J.P. & Zaher, S. (2013). The Demand for Military Spending in Egypt. *Defence and Peace Economics*, 5:3, 231-245, https://doi.org/10.1080/10242694.2013.763454.

[2] Comprehensive State Defence. *Ministry of Defence of Latvia* [Online]. Available at: https://www.mod.gov.lv/en/nozares-politika/comprehensive-state-defence [cited 2023-03-23].
[3] Dudzevičiūtė, G., Bekesiene, S., Meidute-Kavaliauskiene, I. & Ševčenko-Kozlovska, G. (2021). An Assessment of the Relationship between Defence Expenditure and Sustainable Development in the Baltic Countries *Sustainability*.13, no. 12: 6916.
[4] Dunne, J. P., Perlo-Freeman, S. & Soydan, A. (2004). Military Expenditure and Debt in South America. *Defence and Peace Economics*, 15:2, 173-187, https://doi.org/10.1080/1024269032000110540.
[5] Estonia Accepts Finnish Invitation to Go in on Howitzer Tender. *ERR* [Online]. Available at: https://news.err.ee/120567/estonia-accepts-finnish-invitation-to-go-in-on-howitzer-tender [cited 2023-02-13].
[6] Friedrich, P. & Reiljan, J. (2015). Estonian Economic Policy During Global Financial Crises. *EconPol Forum* [Online]. Available at: https://www.ifo.de/DocDL/forum-2015-4-friedrich-reiljan-baltic-tiger-december_0.pdf [cited 2023-03-23].
[7] General Government Debt. *OECD* [Online]. Available at: https://data.oecd.org/gga/general-government-debt.htm [cited 2023-02-16].
[8] Judson, J. (2019). Do the Baltics Need More US Military Support to Deter Russia?. *DefenceNews* [Online]. Available at: https://www.defensenews.com/land/2019/07/15/do-the-baltics-need-more-us-military-support-to-deter-russia/ [cited 2023-03-23].
[9] Latvian Government Confirms Increased Defence Spend [Online]. Available at: https://eng.lsm.lv/article/society/defense/latvian-government-confirms-increased-defense-spend.a450233/ [cited 2023-02-15].
[10] Metha, A. (2018). Latvia Cleared to Buy Black Hawks. *DefenceNews* [Online]. Available at: https://www.defensenews.com/global/europe/2018/08/03/latvia-cleared-to-buy-black-hawks/ [cited 2023-03-23].
[11] Nikolaidou, E. (2008). The Demand for Military Spending: Evidence from the EU15 (1961-2005). *Defence and Peace Economics*, 19:4, 273-292, https://doi.org/10.1080/10242690802166533
[12] Odehnal, J. & Neubauer, J. (2015). Ekonomické determinanty vojenských výdajů – kauzální analýza. *Ekonomický casopis,* 10 (63), 1019-1032. [Online]. Available at: https://www.sav.sk/journals/uploads/0620135310%2015%20Odehnal%20+%20RS.pdf. [cited 2023-03-23].
[13] Pobaltské státy chtějí zvýšit výdaje na armádu na 3 % svých HDP. *Investiční web* [Online]. Available at: https://www.investicniweb.cz/pobaltske-staty-chteji-zvysit-vydaje-na-armadu-na-3-svych-hdp [cited 2023-02-13].
[14] Sezgin, S. & Yildirim, J. (2002). The Demand for Turkish Defence Expenditure. *Defence and Peace Economics*, 13:2, 121-128, https://doi.org/10.1080/10242690210973.
[15] SIPRI Military Expenditure Database. *SIPRI* [Online]. Available at: https://www.sipri.org/databases/milex [cited 2023-02-13].
[16] Skogstad, K. & Compton, R.A. (2021). Country Survey: Canadian Military Expenditure and Defence Policy. *Defence and Peace Economics*, 33:5, 616-636, https://doi.org/10.1080/10242694.2021.1963525.
[17] Smyth, R. & Narayan, P. K. (2009). A Panel Data Analysis of the Military Expenditure – External Debt Nexus: Evidence from Six Middle Eastern Countries. *Journal of Peace Research, 46(2), 235–250.* [Online]. Available at: http://www.jstor.org/stable/25654382 [cited 2023-03-23].
[18] Solarin, S. A. (2018). Determinants of Military Expenditure and the Role of Globalisation in a Cross-country Analysis. *Defence and Peace Economics*, 29:7, 853-870, https://doi.org/10.1080/10242694.2017.1309259.
[19] Yalta, T. A. (2022). The Determinants of Defense Spending in the Gulf Region. *Defence and Peace Economics*, 33:8, 980-992, https://doi.org/10.1080/10242694.2021.1918857.

Investment Portfolio Selection from Shares of Environmental Companies

Juraj Pekár [1], Ivan Brezina [2], Marian Reiff [3]

Abstract. The paper presents the possible use of the portfolio selection optimization model based on the CVaR risk measure when investing in environmental companies. The most important factors in investing are taking risk and return rates into account. Currently, investing in environmental funds is a popular tool. When creating a portfolio, it is possible to make an investment by choosing an investment strategy provided by a financial institution, but it is also possible to create own portfolio. In the contribution, we point out the possible use of the portfolio selection model based on the CVaR risk rate when creating own composition of assets. In the paper, we consider investing in the largest environmental companies selected by Value.Today analytics company.

Keywords: return, CVaR, environmental investment

JEL Classification: G11
AMS Classification: 91B30, 90C90

1 Introduction

The pressure to protect the environment is constantly growing, and more and more companies are focusing on environmental sustainability. Since the corporate strategy is not focused on environmental activities just as a marketing move but also becomes a real strategy, with which investment is also connected. In the end, meeting ecological requirements is also associated with increased costs. Therefore, it requires investments if companies will function and bring more ecological and socially responsible solutions. Therefore, investors' interest in ecologically or socially responsible companies is growing [1], [2].

Investing in environmental (green, ecological) companies has been one of the tools for the sustainable development of society since the 1990s. Environmental investing is generally the search for investment opportunities that benefit the environment. The goal of environmental investment is to support business activities that have a positive impact on the natural environment.

The paper presents the possibility of using the classic analysis of the returns of effective portfolios constructed based on the optimization model of portfolio selection based on the CVaR risk measure in environmental investment. The shares of the 19 largest global environmental companies selected by analytics company Value.Today were selected for analysis.

Company Value.Today is a software analytics company that provides corporate information, company financials, and global financial news. The Value.Today's concept is primarily based on providing information related to business analyses and the characteristics of how different companies in different sectors operate in different countries worldwide. The company provides data on global markets, sector performance, company market value, annual and quarterly company results, balance sheet and cash flow data, and key analytical indicators related to corporate companies.

Because the authors wanted to analyze investment strategies based on a portfolio selection model using all available data, the analysis was performed on daily historical data from 19.5.2017 to 30.12.2022 based on Value.Today data for selected stocks of the 19 largest global environmental companies [13], [15].

[1] University of Economics in Bratislava/Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, juraj.pekar@euba.sk.
[2] University of Economics in Bratislava/Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, ivan.brezina@euba.sk.
[3] University of Economics in Bratislava/Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, marian.reiff@euba.sk.

2 Risk Measure

In the paper, the Conditional Value at Risk (CVaR) risk measure is applied, because is currently one of the most used risk measures (other risk measures can be lower semi-variance, Value at Risk [4], [5] and [8], but also risk measures based on the drawdown principle (the drawdown function represents the difference between the maximum return of the portfolio up to time T and the value of the portfolio at time T) absolute Drawdown, maximum Drawdown, average Drawdown, Drawdown at risk, or conditional Drawdown at risk [2].

The CVaR risk measure is defined as the expected loss exceeding the Value at Risk (VaR - an indicator that gives an estimate of the highest decline in the value of an asset or portfolio of assets over the monitored period that can be expected under normal conditions) (Konno and Yamazaki 1991). Based on this definition, CVaR only considers loss values higher than the VaR value. The CVaR value is then defined:

$$CVaR_\alpha(X) = E\left(L(X) \middle| L(X) \geq VaR_\alpha\right) \tag{1}$$

where X denotes the random variable representing the return, $L(X) = -X$ denotes the loss function of the random variable X and is the value at risk at the significance level α (the worst possible loss in the observed time interval that can be expected at $100\alpha\%$).

Assuming the existence of a discrete random variable X, represented by the return vector $\mathbf{r} = (r_1, r_2, \ldots r_T)$, where T is the number of components and p_t represents the significance of observation for $t = 1, 2, \ldots T$. The above risk measure can be defined:

$$CVaR_\alpha(X) = VaR_\alpha + \frac{1}{\alpha}\sum_{t=1}^{T} p_t \max(-(r_t + VaR_\alpha), 0) \tag{2}$$

Rockafellar and Uryasev [12] demonstrated that linear programming tools could be used to optimize CVaR risk measures. Other published studies present the fact that risk optimization using the CVaR risk measure can be implemented for large portfolios and a large number of scenarios with relatively small computing resources [8], [9], [12] and [14].

3 Portfolio Selection Model Based on CVaR

In general, let's consider the creation of a portfolio consisting of n assets (bank deposits, shares, bonds, or mutual funds) with return vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_n$, which represent discrete random variables, while in the search for the optimal portfolio, a measure was chosen CVaR risks. Furthermore, the constructed model presents a possible portfolio optimization approach, i.e., a procedure for diversifying assets into a portfolio so that the risk is minimal for a given return [6], [7], [10], [11] and [16].

Let $E(\mathbf{r}_j)$ represent the expected return of the j-th asset (also denote E_j as the expected return of the portfolio) and r_{jt} is the t-th component ($t = 1, 2, \ldots T$) of the discrete random variable X_j represented by the vector of returns r_j for $j = 1,2,\ldots n$. It can be assumed that the investor invests in individual assets with a certain share w_j (the share of the j-th asset in the total investment), which represents the vector of weights $\mathbf{w} = (w_1, w_2, \ldots w_n)^{\mathrm{T}}$. Then the expected return of the portfolio can be determined as $\sum_{j=1}^{n} w_j E(\mathbf{r}_j)$. Furthermore, the assumption can be formulated that the expected value of the random variable X_j can be expressed as an average calculated from these data. When constructing the model, we will use the variables $z_t \geq 0$ ($t = 1, 2, \ldots T$), which take on the value of the difference between VaR and the return of the portfolio in state t, if the return is lower than VaR, or will be equal to zero based on the level of significance α.

Since the objective of the optimization is to minimize the risk in the CVaR space at a given value of the expected return of the portfolio, the minimization task for the CVaR has the form:

$$\begin{aligned}
&\min\left\{VaR_\alpha + \frac{1}{\alpha}\sum_{t=1}^{T} p_t z_t\right\} \\
&z_t + \sum_{j=1}^{n} r_{jt} w_j + VaR_\alpha \geq 0, t = 1,\ 2, \ldots T \\
&\sum_{j=1}^{n} E_j w_j \geq E_P \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_1, w_2, \ldots w_n \geq 0, z_1, z_2, \ldots z_T \geq 0
\end{aligned} \tag{3}$$

Model (3) was applied to determine efficient portfolios for different values of the required minimum EP return. At the same time, the result was calculated investment weights in individual assets while minimizing the CVaR risk function.

4 Investment Recommendation Based on the Portfolio Selection Model

In general, an investment portfolio represents a set of various financial instruments that an investor owns with the expectation and purpose of valuing his funds. For example, creating an environmental (ecological, green) portfolio means placing funds in the stock market by buying shares of companies that focus on maintaining the environment's safety.

Based on Value.Today data, the shares of the 19 largest global environmental companies were selected for analysis (of course, other analytical entities could select a different set of the most significant environmental companies based on their criteria):

Waste Management (WM), Republic Services (RSG), Waste Connections (WNC), Veolia Environnement (VEOEY), China Conch Venture Holdings (0586.HK), Stericycle (SRCL), Clean Harbors (CLH), Fomento De Construcciones Y Contratas (FCC.MC), China Everbright International (CNE.SG), Cleanaway Waste Management (CWY.AX), Companhia De Saneamento Basico Do Estado De Sao Paulo – Sabesp (SAJA.BE), Casella Waste Systems (CWST), Beijing Enterprises Water Group (0371.HK), Daiseki (9793.T), Munters Group Ab (MTRS.ST), Harsco Corporation (HSC), Metawater (9551.T), China Everbright Greentech (CK7.F), Dredging Corporation Of India (DREDGECORP.NS- in the table marked as DRE). The analysis was carried out on daily historical data for the period from 19.5.2017 to 30.12.2022.

A mathematical programming task (3) was solved to create an environmental portfolio based on considered historical data for the 19 listed actions (3). By solving the problem, efficient portfolios can be obtained for the specified values of the expected weekly returns shown in Table 1 in the column labeled EP. The stated values are obtained as the smallest and largest values of the portfolio's expected returns, while the other values are determined by dividing the interval into equal parts.

CVaR	E_p	**0586.HK**	**9551.T**	**9793.T**	**CK7.F**	**CNE.SG**	**CWST**	**CWY.AX**	**DRE**	**FCC.MC**	**SAJA.BE**	**WM**
5.247	0.349	0.0	24.8	0.0	7.2	3.4	9.7	30.6	2.2	7.5	2.4	12.4
5.393	0.431	0.0	19.0	1.2	3.7	8.5	15.2	34.2	0.0	0.0	0.2	18.0
5.716	0.514	0.0	20.1	4.0	0.0	17.1	24.8	34.0	0.0	0.0	0.0	0.0
6.141	0.596	0.3	10.2	3.7	0.0	22.7	32.6	25.3	0.0	0.0	0.0	5.3
6.581	0.678	0.0	0.0	4.2	0.0	27.6	41.3	19.2	0.0	0.0	0.0	7.8
7.078	0.761	0.0	0.0	5.3	0.0	37.3	47.2	10.2	0.0	0.0	0.0	0.0
7.732	0.843	0.0	0.0	0.2	0.0	49.2	50.6	0.0	0.0	0.0	0.0	0.0
8.876	0.925	0.0	0.0	0.0	0.0	74.5	25.5	0.0	0.0	0.0	0.0	0.0
11.03	1.00	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1 Distribution of investment in effective portfolios (unlisted shares have weights equal to 0, thus not invested in any of the analyzed periods). All values in the table have the units %. Source: Author's own calculations

Table 1 shows the calculated solutions. First, the objective function's value representing the minimum CVaR value is given in the column labeled CVaR. The second column (Ep) presents the portfolio's expected return values. In other columns, the shares invested in individual shares are listed at different expected return values. It is clear from Table 1 that the recommendation based on the portfolio selection model using the input data is the investment in shares 0586.HK, 9551.T, 9793.T CK7.F, CNE.SG, CWST, CWY.AX, DREDGECORP.NS, FCC.MC, SAJA.BE, WM, which will form the investment portfolio.

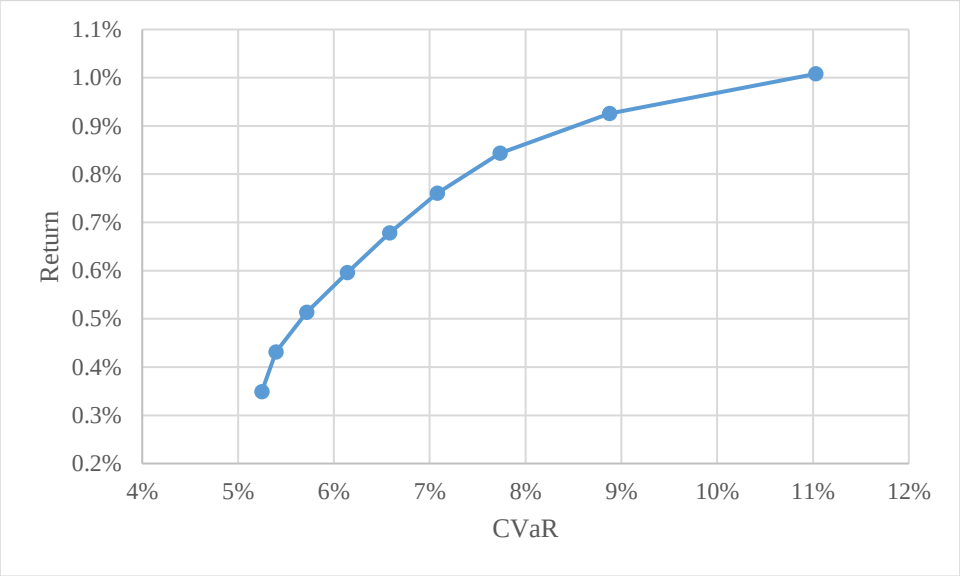


Figure 1 Effective frontier. Source: Author's own calculations

Figure 1 shows a set of efficient portfolios that represent the relevant risk values at the specified expected return values in the case of investment in the listed environmental companies.

5 Conclusion

Based on the assumption that we are considering investing in environmental companies, investors should have the tools to decide which assets can and should be invested in with the view of the most significant profit with the least risk. Therefore, every investor should be interested in alternative ways of investing or o different investment models. Investment companies offer different forms of investment, for example, in an environmental stock index or a portfolio of environmental stocks created by an investment company. When placing financial resources in environmental investments, the investor should respect certain principles to ensure that his investments align with the values of the company's environmental goals. Investing in "green" companies is generally riskier than investing in other equity strategies. This fact is determined by the fact that many ecological companies are in the development phase, with low revenues and a high valuation of revenues. However, if supporting environmental businesses is important to investors, then this type of investment can be an attractive way to invest available investment funds. The subject of the presented analysis was a possible approach to the effective application of the classical analysis of the returns of effective portfolios constructed based on the optimization model of portfolio selection based on the CVaR risk measure in environmental investment. The shares of the 19 largest global environmental companies selected by Value.Today were selected for analysis.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0120/23 „Environmental models as a tool for ecological and economic decisions making “.

References

[1] Azhgaliyeva, D., Kapsalyamova, Z. & Low, L. (2019). Implications of Fiscal and Financial Policies on Unlocking Green Finance and Green Investment. *In Handbook of green finance* (pp. 427-457). Springer, Singapore.

[2] Cheklov, A., Uryasev, S. & Zabarankin, M. (2004). Portfolio optimization with drawdown constraints. In *Supply chain and finance*, 209-228. https://doi.org/10.1142/9789812562586_0013.
[3] Eyraud, L., Clements, B. & Wane, A. (2013). Green investment: Trends and determinants. *Energy Policy* (60), 852-865. https://doi.org/10.1016/j.enpol.2013.04.039.
[4] Kang, Z., Li, X. & Li, Z. (2020). Mean-CVaR portfolio selection model with ambiguity in distribution and attitude. *Journal of Industrial & Management Optimization*, *16*(6), 3065-3081. https://doi.org/10.3934/jimo.2019094.
[5] Konno, H. & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. In *Managment science*. *37*(5), 519-53. https://doi.org/10.1287/mnsc.37.5.519.
[6] Krokhmal, P., Uryasev, T. & Palmquist, J. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, *4*, 43-68.
[7] Liu, N., Chen, Y. & Liu, Y. (2018). Optimizing portfolio selection problems under credibilistic CVaR criterion. *Journal of Intelligent & Fuzzy Systems*, *34*(1), 335-347. https://doi.org/10.3233/JIFS-171298.
[8] Pekár, J. (2015). *Modely matematického programovania na výber portfólia*. Bratislava: Vydavateľstvo EKONÓM.
[9] Pekár, J., Brezina, I. & Brezina, I. jr. (2018). Portfolio Selection Model Based on CVaR Performance Measure. *Quantitative Methods in Economics: Multiple Criteria Decision Making XIX*. Bratislava: Letra Edu.
[10] Pekár, J., Brezina, I. & Reiff, M. (2022a). *Modely výberu portfólia*. Bratislava: Vydavateľstvo EKONÓM.
[11] Pekár, J., Brezina, I. &Reiff, M. (2022b). Determining the Investor's Strategy During the COVID-19 Crisis Based on CVAR Risk Measure. *Strategic Management and Decision Support Systems in Strategic Management SM 2022*. Subotica: University of Novi Sad. https://doi.org/10.5937/StraMan2200029P.
[12] Rockafellar, R. T. & Uryasev, S. (2002). Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*. *26*(7), 1443−1471. https://doi.org/10.1016/S0378-4266(02)00271-6.
[13] Source of Data [Online]. Available at: https://finance.yahoo.com [cited 2023-04-20].
[14] Sun, Y. F, Aw, E. L. G., Li, B., Teo, K. L. & Sun, J. (2020). Cvar-based robust models for portfolio selection. *Journal of industrial and management optimization*, *16*(4), 1861–1871. https://doi.org/10.3934/jimo.2019032.
[15] World Top Companies [Online]. Available at: https://www.value.today/ [cited 2023-04-20].
[16] Xu, Q., Zhou, Y., Jiang, C., Yu, K. & Niu, X. (2016). A large CVaR-based portfolio selection model with weight constraints. *Economic Modelling*, *59*, 436-447. https://doi.org/10.1016/j.econmod.2016.08.014.

Application of CCR and SBM Models in Measuring the Efficiency of IT Clusters in the Czech Republic and Slovakia

Natalie Pelloneová[1], Vladimíra Hovorková Valentová[2]

Abstract. This paper analyses the technical efficiency of members of two cluster organizations operating in the Czech Republic and Slovakia. There are many methods of measuring the efficiency of companies. This paper focuses on the Charnes, Cooper, Rhodes, and Slack Based Measure models, comparing the efficiency analysis results calculated by the two aforementioned methods. The presented research examines four research samples consisting of member companies in the IT Cluster and Košice IT Valley in 2013 and 2020. Both clusters were created as a result of a cluster initiative and brought together companies from the ITC sector. The firms that form the core of the above-mentioned clusters are active in sectors with the following statistical classifications: NACE 620100, 620200, and 620900. This paper compares the results obtained using both methods.

Keywords: IT sector, cluster organization, data envelopment analysis, CCR model, Slack Based Measure

JEL Classification: C10, L86, C67, C44
AMS Classification: 90B90, 90C90

1 Introduction

The concept of clustering is one of the critical sources of growth in modern market economies. Companies perceive cluster organizations as a very attractive way of cooperation. If firms become members of a suitable cluster organization, they have a great opportunity to improve their competitive position and financial or innovation performance [14], [15]. In the following research, we draw on Porter's [14] definition of a cluster: "A cluster is a geographic concentration of interrelated companies and affiliated institutions in a particular industry that is communally connected and complementary." According to this definition, it is clear that the firms in the cluster are interconnected and complementary. The linkages between enterprises can be vertical (supply-customer chain) or horizontal (complementary products and services, use of similar inputs, technologies, labor, etc.). The second key feature of clusters, according to Porter's definition, is geographical proximity. Collocation increases the benefits of networks of direct and indirect interactions between firms [1]. There is much debate about the economic benefits of clusters, and experts generally agree that clusters bring various benefits to the firms involved [7]. The prevailing scholarly studies focus mainly on the economic benefits of clusters and innovation, e.g. [5], [12], and [10], on improving competitiveness, e.g. [2], and on improving financial performance, e.g. [8], [11], and [16]. Clusters provide member firms with access to strategically important resources, reduce transportation costs, and facilitate access to customers and labor [14]. According to Jirčíková et al. [9], other benefits gained by member firms include competitive cost structure, gain of new financial resources, productivity growth, and performance growth. In the Czech literature, despite a considerable amount of research, there is no comprehensive methodology focused on the impact of firms' involvement in a cluster on their efficiency. The present paper is thus one of the few that deals with the relationship between efficiency and firm membership in cluster organizations and builds on the authors' previous research [19], [20], and [13]. Efficiency analysis is essential for assessing the financial performance of member firms. A firm is considered efficient if it is able to generate maximum returns by efficient use of its financial resources. There are many methods in the literature that deal with measuring the efficiency of firms. This paper uses the Charnes, Cooper, and Rhodes (hereafter CCR) and Slack Based Measure (hereafter SBM) models introduced by Charnes et al. [4] and Tone [17]. The research uses secondary data obtained from the Czech MagnusWeb database and the Slovak Register of Financial Statements. The research sets two main objectives, which are to measure the technical efficiency of member firms in IT cluster organizations and also to compare the efficiency scores between the CCR and SBM models. For this purpose, the research evaluates two cluster organizations in the Czech Republic and Slovakia: the IT Cluster (CZ) and Košice IT Valley (SK). The present research focuses on these countries for three reasons. First, these

[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, vladimira.valentova@tul.cz.

countries are economically comparable. Second, both countries have recently supported the emergence of clusters through national policies and significant public investment. Third, the high availability of data and financial reporting in these countries.

2 Methods and Data

This chapter describes the selected methods used in the presented research, the aim of which was to analyze the technical efficiency of member companies of the IT Cluster and Košice IT Valley between 2013 and 2020. Data for 2021 and later are not yet available for a significant proportion of member businesses. Older data, on the other hand, are not available for Slovak businesses. For the purposes of this paper (testing phase), only two years were evaluated (the initial year, i.e., 2013, and the final year 2020). Further research envisages extending the research period by additional years. In order to measure efficiency, the method of data envelopment analysis was used, which is designed to evaluate the efficiency of homogeneous production units (in this case, member firms) based on selected input and output variables.

2.1 Data Envelopment Analysis

The methods used to evaluate technical efficiency can be distinguished into parametric and non-parametric. The most widely used non-parametric methods include data envelopment analysis (hereafter DEA). The data envelopment analysis method is prevalent among experts in measuring the efficiency of companies. DEA is designed to evaluate the efficiency of production units that are homogeneous. The first model used is the input-oriented Charnes, Cooper, and Rhodes model assuming constant returns to scale. Its mathematical formulation is as follows:

$$\begin{gathered} min\ \theta_q \\ s.t. \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \quad i = 1,2,\dots,m, \\ \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \quad k = 1,2,\dots,r, \\ \lambda_j \geq 0, \ \ j = 1,2,\dots,n, \end{gathered} \tag{1}$$

Where λ_j, j = 1, 2, ..., n are weights of all DMUs, s_i^-, i = 1, 2, ..., m and s_k^+, k = 1, 2, ..., r are slack/surplus variables, θ_q is the efficiency score of the DMUq [4]. The second model used is the SBM model authored by Tone [17]. This model measures efficiency using additive variables. The value of the additive variables expresses the distance of the unit under study from the production possibilities frontier calculated by the model. The input-oriented SBM model has the following form [18]:

$$\begin{gathered} \min \tau_{j_0} = \alpha - \frac{1}{W}\sum_{k=1}^{W}\frac{s_k^-}{x_{k_{j0}}} \\ \alpha + \frac{1}{t}\sum_{r=1}^{t}\frac{s_r^+}{y_{r_{j0}}} = 1, \\ \sum_{j=1}^{n} x_{k_j}\lambda_j + s_k^- = \alpha x_{kj_0}, \quad k = 1,\dots,w, \\ \sum_{j=1}^{n} y_{r_j}\lambda_j - s_r^+ = \alpha y_{rj_0}, \quad r = 1,\dots,t, \\ s.t.\ \lambda_j \geq 0, j = 1,\dots n, s_k^- \geq 0, k = 1,\dots,w, s_r^+ \geq 0, r = 1,\dots,t\ \alpha > 0. \end{gathered} \tag{2}$$

Where τ_{j_0} is the efficiency measure of the unit under study. Efficient units have an efficiency measure of τ_{j_0} = 1, i.e., all additive variables are equal to 0. Inefficient units have an efficiency measure τ_{j_0} less than 1 [18]. All calculations were performed in MaxDEA Ultra software. Efficiency results were generated for both the CCR-I

and SBM models. Efficiency is measured on a scale of 0 to 1, where a value of 1 indicates that the unit (i.e., firm) is efficient, and a value less than 1 indicates that the unit (i.e., firm) is inefficient.

2.2 Data and Variables

In order to get a better overview of the evaluated cluster organizations, this section of the paper presents their essential characteristics and selected data. The first cluster studied is the Czech IT Cluster. The cluster brings together business entities and educational institutions active in ICT in the Moravian-Silesian Region. The IT Cluster was established in 2006 and brings together member entities with statistical classification CZ-NACE 58290, 620100, 620200, and 620900. In the analyzed period, the IT Cluster had a total of 21 members. The second cluster analyzed is the Slovak cluster Košice IT Valley. The cluster has a critical role in developing the ITC sector in the entire Košice region. The cluster was established in 2007 and brings together member entities with statistical classification SK-NACE 62010, 62030, 62090, and 63110. In the analyzed period, the Košice IT Valley cluster had 55 members. In the research, only firms that have been members of a cluster organization for the same length of time can be compared; only these firms can be considered the cluster's core. The core of the IT Cluster consists of a total of 12 business entities. In the case of Košice IT Valley, the core consists of 17 companies.

Calculations were performed for four research datasets. All research datasets contained two inputs (total assets and fixed assets) and two outputs (sales and profit after tax; henceforth, EAT). All inputs and outputs for both cluster organizations in 2013 and 2020 are shown in Table 1 and Table 2.

Company	Assets (thous. CZK)		Long-term invested resources (thous. CZK)		Revenues (thous. CZK)		EAT (thous. CZK)	
	2013	2020	2013	2020	2013	2020	2013	2020
Kvados	129,369	137,790	97,941	65,787	122,894	142,065	9,794	3,125
Iteuro	13,196	33,049	5,450	5,062	29,074	0	1,653	0
NAM systém	100,109	298,828	91,270	287,211	87,493	151,798	6,000	40,810
Fullcom systems	8,959	12,771	5,618	6,912	12,309	22,068	227	171
D3Soft	32,511	19,169	16,193	13,554	46,751	45,840	1,086	934
Techniserv IT	104,821	157,598	36,214	58,208	209,930	338,386	4,726	16,363
NITTA Systems	868	5,263	551	2,661	0	6,990	236	1,164
XEVOS Solutions	678	60,234	155	37,503	92	136,685	265	13,536
Vítkovice IT solutions	442,198	228,252	233,224	124,112	635,133	303,063	86,992	4,980
K2 atmitec	170,304	128,463	94,947	96,885	186,005	263,304	10,176	25,891
Eltodo	1,403,760	1,489,370	805,713	779,361	1,176,096	1,469,666	153,931	53,811
Pronix	49063	182,650	17,536	91,086	115,462	262,700	1,207	30,066
Tietoevry Connect Czechia	459,520	517,526	162,721	178,988	1,790,655	2,843,898	156,145	118,068

Table 1 Input and output variables of the IT Cluster (CZ)

Company	Assets (EUR)		Long-term invested resources (EUR)		Revenues (EUR)		EAT (EUR)	
	2013	2020	2013	2020	2013	2020	2013	2020
Asbis SK	66,452,937	75,402,024	5,278,151	7,008,137	211,052,334	257,625,691	889,219	896,621
bart.sk	278,909	1,129,809	207,443	897,043	578,344	2,364,490	22,007	488,598
Cisco Systems Slovakia	4,209,569	6,195,525	3,415,740	5,256,477	4,991,809	5,361,948	295,568	208,005
Deloitte Audit	6,167,682	4,994,669	1,467,507	1,354,377	12,943,438	12,254,757	512,511	175,657
Elcom	9,272,371	9,132,019	6,051,746	7,613,686	9,755,443	4,109,952	328,969	-794,929
Esten	322,205	525,757	72,718	103,114	926,573	1,205,679	60,519	71,670
Fpt Slovakia	3,981,153	3,431,718	843,211	1,584,426	18,655,207	13,020,728	801,397	1,503,588

gd - Team	1,383,782	2,115,762	366,237	811,173	2,557,233	7,478,236	9,162	115,112
Matsuko	297,816	759,351	65,839	759,351	536,574	0	2,470	7,896
Promiseo	72,547	732,155	56,711	227,475	177,207	1,269,948	27,148	79,468
SAP Slovensko	25,781,048	35,679,324	15,417,617	20,505,741	39,023,783	46,697,494	3,021,602	2,104,284
Software AG Development Center Slovakia	613,169	875,493	417,213	618,893	1,465,784	2,111,142	56,795	125,221
Telegrafia	8,962,624	10,431,239	2,881,962	4,833,316	5,911,686	5,691,720	455,006	627,667
TORY Consulting	4,571,785	5,670,969	1,718,848	1,053,833	6,569,562	6,802,126	488,333	348,823
T-Systems Slovakia	49,316,884	58,563,136	27,084,532	33,041,695	94,278,785	142,089,977	3,353,262	5,375,871
Unique People	37,713	989,531	18,863	576,016	14,071	2,883,751	17	383,097
VSL Software	3,788,685	3,202,717	1,814,904	2,605,495	2,583,402	3,732,819	346,464	985,930

Table 2 Input and output variables of the Košice IT Valley cluster (SK)

3 Research Results

Table 3 shows the technical efficiency measures obtained using the CCR-I and SBM model for the case of the Czech IT Cluster. Using these two models, Tietoevry Connect Czechia and XEVOS Solutions were identified as efficient in 2013. The remaining ten companies are already quite far behind in this characteristic. Both models show the same result where two firms are considered efficient while the other ten firms are considered inefficient. However, the SBM model provides a more detailed and accurate result of inefficiency assessment in terms of excess inputs and lack of outputs. In 2020, only one company, Tietoevry Connect Czechia, was identified as efficient by the CCR-I model. This company was also identified as efficient in terms of the SBM model. CCR-I identified the remaining 11 companies as inefficient. In addition, the SBM model has been identified as efficient by NITTA Systems. This company also achieved a relatively good technical efficiency score according to the CCR-I model. The last row of Table 3 shows the median value. The resulting median values confirm the hypothesis that the efficiency measure obtained by the SBM model is mostly lower than that obtained by the CCR-I model. It is also clear from the median that the Czech cluster achieved better technical efficiency scores in 2020 compared to 2013 (for both models).

DMU names	**2013**		**2020**	
	CCR-I	**SBM**	**CCR-I**	**SBM**
D3Soft	0.369	0.132	0.435	0.239
ELTODO	0.308	0.209	0.180	0.140
FULLCOM systems	0.353	0.095	0.314	0.092
K2 atmitec	0.280	0.177	0.883	0.391
Kvados	0.244	0.171	0.188	0.114
NAM systém	0.224	0.137	0.599	0.110
NITTA Systems	0.696	0.017	0.969	**1.000**
PRONIX	0.604	0.128	0.722	0.330
TECHNISERV IT	0.527	0.213	0.455	0.413
Tietoevry Connect Czechia	**1.000**	**1.000**	**1.000**	**1.000**
VÍTKOVICE IT SOLUTIONS	0.551	0.376	0.242	0.113
XEVOS Solutions	**1.000**	**1.000**	0.985	0.479
Median	**0.448**	**0.174**	**0.527**	**0.285**

Table 3 Results for IT Cluster (CZ)

Table 4 shows the technical efficiency measures obtained using the CCR-I and SBM model for the case of the Slovak Košice IT Valley. Using the CCR-I model, Asbis SK, Fpt Slovakia, and Promiseo came out as efficient in 2013. The remaining 14 companies lag behind in this characteristic. In 2013, the SBM model identified Asbis

SK, Esten, Fpt Slovakia, Promiseo, SAP Slovakia, T-Systems Slovakia, and Unique People as efficient. Fewer companies were labeled with the CCR-I model in 2020 than in 2013 (only Asbis SK and Fpt Slovakia). The SBM model was also identified by fewer companies as efficient (Asbis SK, bart.sk, Esten, Fpt Slovakia, T-Systems Slovakia, and Unique People). The last row of Table 2 shows the median value. The resulting median values confirm the hypothesis that the efficiency measure obtained by the SBM model is mostly lower than that obtained by the CCR-I model. It is also clear from the median that the Slovak cluster achieved better technical efficiency scores in 2020 compared to 2013 (for both models).

DMU names	**2013**		**2020**	
	CCR-I	**SBM**	**CCR-I**	**SBM**
Asbis SK	**1.000**	**1.000**	**1.000**	**1.000**
bart.sk	0.443	0.290	0.987	**1.000**
Cisco Systems Slovakia	0.287	0.185	0.228	0.089
Deloitte Audit	0.448	0.406	0.682	0.205
Elcom	0.225	0.131	0.119	0.056
Esten	0.914	**1.000**	0.870	**1.000**
Fpt Slovakia	**1.000**	**1.000**	**1.000**	**1.000**
gd - Team	0.394	0.055	0.952	0.269
Matsuko	0.384	0.086	0.024	0.000
Promiseo	**1.000**	**1.000**	0.477	0.505
SAP Slovensko	0.416	**1.000**	0.345	0.397
Software AG Development Center Slovakia	0.510	0.330	0.636	0.456
Telegrafia	0.216	0.150	0.144	0.140
TORY Consulting	0.425	0.304	0.436	0.344
T-Systems Slovakia	0.408	**1.000**	0.639	**1.000**
Unique People	0.080	**1.000**	0.884	**1.000**
VSL Software	0.324	0.159	0.703	0.336
Median	**0.416**	**0.330**	**0.639**	**0.397**

Table 4 Results for Košice IT Valley (SK)

4 Conclusion

This paper uses the Slack-based Measure (SBM) method to measure the efficiency of clustered firms compared to the traditional Charnes, Cooper, and Rhodes (CCR-I) model. These methods were applied to data from 2013 and 2020 for two clusters - IT Cluster and Košice IT Valley. From the outputs obtained and presented in the previous chapter, it is clear that in 2013, only two firms in the IT Cluster were identified as technically efficient through the two models used (Tietoevry Connect Czechia and XEVOS Solutions). In 2020, both models identified only one company as efficient, Tietoevry Connect Czechia. In addition, the SBM model identified NITTA Systems as an efficient firm, and it does not score poorly in the CCR model either (0.969). It can be seen from the median values of the calculated scores from both years that the value in 2020 is higher than in 2013. If we look at the results for Košice IT Valley, we see that both models were identified as efficient in 2013 by three companies (Asbis SK, Fpt Slovakia, and Promiseo). Looking at the results separately, the CCR model identified three firms as efficient, whereas the SBM model identified seven firms. In 2020, only two units (Asbis SK and Fpt Slovakia) identified both models as efficient. The BSM model alone was considered efficient by a total of six companies. The median scores in 2013 are lower for both models than in 2020.

It is evident that although the median values could suggest an increase in the technical efficiency measure for both clusters analyzed, in the aggregate, there has been a decrease in the number of firms that consider themselves technically efficient. The SBM model alone has seen two technically efficient firms in 2013 and an IT Cluster in 2020. This finding thus joins studies that do not confirm the positive impact of cluster membership [3], and [6].

References

[1] Asheim, B. T., Cooke, P. & Martin, R. (2006). *Clusters and regional development: critical reflections and explorations*. New York: Routledge, Taylor.

[2] Bialic-Davendra, M., Pavelková, D. & Vejmělková, E. (2014). *The clusters phenomenon in the selected Central European countries*. Newcastle upon Tyne: Cambridge Scholars.

[3] Bresnahan, T., Gambardella, A. & Saxenian A. (2001). 'Old Economy' Inputs for 'New Economy' Outcomes: Cluster Formation in the New Silicon Valleys. *Industrial and Corporate Change*, *10*(4), 835–860.

[4] Charnes, A., Cooper, W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, *2*(4), 228–339.

[5] Dhewanto, W., Prasetio, E. A., Ratnaningtyas, S., Herliana, S., Chaerudin, R., Aina, Q., Bayuningrat H., R. & Rachmawaty, E. (2012). Moderating effect of cluster on Firm's innovation capability and Business Performance: *A conceptual framework. Procedia - Social and Behavioral Sciences*, *65*, 867–872.

[6] Enright, M. J. (2003). Regional Clusters: What We Know and What We Should Know. In J. Bröcker, D. Dohse & R. Soltwedel (Eds.), *Innovation Clusters and Interregional Competition* (pp. 99–129). Berlin: Springer.

[7] Eriksson, S. (2009). *Clusters: A Survey of Research within Localized Economic Growth*. Jönköping: Jönköping International Business School.

[8] Harrison, B. (1997). *Lean and mean: The changing landscape of economic power in the age of flexibility*. New York: Guilford Press.

[9] Jirčíková, E., Remeš, D. & Pavelková, D. (2006). Zvyšování výkonnosti podniků prostřednictvím zapojení do průmyslových klastrů. In *Mezinárodné vedecké dni 2006 – zborník príspevkov z medzinárodnej vedeckej konferencie na téma „Konkurencieschopnosť v EU – výzva pre krajiny V4"* (pp. 164–170). Nitra: Slovenská poľnohospodárska univerzita v Nitre.

[10] Khan, J. H. & Ghani, J. A. (2004). Clusters and entrepreneurship: implications for innovation in a developing economy. *Journal of Developmental Entrepreneurship*, *9*(3), 221–38.

[11] Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, *99*(3), 483–499.

[12] Lecocq, C., Leten B., Kusters, J. & van Looy, B. (2011). Do firms benefit from being present in multiple technology clusters? An assessment of the technological performance of biopharmaceutical firms. *Regional Studies*, *46*(9), 1107–1119.

[13] Pavelková, D., Žižka, M., Homolka, L., Knapková, A. & Pelloneová, N. (2021). Do clustered firms outperform the non-clustered? Evidence of financial performance in Traditional Industries. *Economic Research-Ekonomska Istraživanja*, *34*(1), 3270–3292.

[14] Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

[15] Porter, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, *76*(6), 77–90.

[16] Storper, M. (1995). Regional technology coalitions: An essential dimension of national technology policy. *Research Policy*, *24*(6), 895–911.

[17] Tone, K. (2001). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, *130*(2), 498–509.

[18] Zýková, P. & Jablonský, J. (2018). Analýza efektivnosti obchodních řetězců v České republice. *Trendy v Podnikání*, *8*(3).

[19] Žižka, M., Valentová, V. H., Pelloneová, N. & Štichhauerová, E. (2018). The effect of clusters on the innovation performance of enterprises: Traditional vs new industries. *Entrepreneurship and Sustainability Issues*, *5*(4), 780–794.

[20] Žižka, M. & Pelloneová, N. (2019). Do clusters with public support perform better? Case study of Czech cluster organizations. *Administratie Si Management Public*, *1*(33), 20–33.

Strong Generalized Eigenvector of Fuzzy EA-Interval Matrices

Ján Plavka[1]

Abstract. Generalized eigenproblems in a max-min (fuzzy) algebra (for given matrices A, B find a vector x and a constant λ such that $A \otimes x = \lambda \otimes B \otimes x$, where the standard pair of operations, plus and times, have been replaced by the operations $\oplus$ = max and $\otimes$ = min) plays important role in practical problems related to scheduling optimization, modeling of fuzzy discrete event dynamic systems and fuzzy analysis. Steady state of synchronization discrete event dynamic system can be characterized by a solution of a generalized eigenproblem. In practice, the entries of matrices and vectors are considered as inexact data (intervals). An interval vector $\boldsymbol{X}$ is said to be a strong eigenvector of square interval matrices $\boldsymbol{A}, \boldsymbol{B}$ if $A \otimes x = \lambda \otimes B \otimes x$ holds for each x in $\boldsymbol{X}$, A in $\boldsymbol{A}$, B in $\boldsymbol{B}$ and for some constant λ. We suppose that an interval matrix $\boldsymbol{A}$ can be split into two subsets according to exists–forall quantifications of its interval entries (EA-interval matrix). The properties of strong generalized eigenvector of EA-interval matrices are studied and characterizations of equivalent conditions are presented. As a consequence of the obtained results, efficient algorithm for checking obtained equivalent conditions is introduced.

Keywords: matrix, interval, eigenvector

JEL Classification: C60
AMS Classification: 08A72, 90B35, 90C47

1 Motivation, Preliminaries and Basic Definitions

Max-min (fuzzy) algebra is algebraic structure where the addition and the multiplication have been formally replaced by the operations of maximum and minimum. Max-min (fuzzy) algebra can be used in a range of practical problems related to scheduling, optimization and also in the modeling in economy. As usual, two arithmetical operations are naturally extended to matrices and vectors.

The investigation throughout this paper will be motivated by the slightly adapted example presented in [2].

Consider a supply control system of commodities consists of n supply points s_i and one control authority point C. The commodities are sent from s_i, $i \in N$ to the control authority point C whereby the checked up commodities have to be sent back to s_i. Moreover, assume that the connection between s_i and C is only possible via one of n available control analyzers r_j, $j \in N$. Further assume that the connections between the s_i and the r_j are one-way connections, and that the capacity of the connection between s_i and r_j is equal to a_{ij}. The security analyzers r_j are connected with C by two-way connections with capacities x_j in both directions. The commodities are sent in commodity packets, and any commodity packet is sent over just one connection as an inseparable unit. Thus, the total capacity of the connection between s_i and C is $\max_{j\in N}\{\min\{a_{ij}, x_j\}\}$, whereby different used connections are comprised as the maximum of capacities (not as their sum).
The transport of commodities from C to s_i is made over other one-way connections between the control analyzer r_j and s_i with capacities equal to b_{ij}. Since the connections between C and r_j are two-way connections, the total capacity of the connection between C and s_i is equal to $\max_{j\in N}\{\min\{b_{ij}, x_j\}\}$. The task is to achieve a synchronization process in which the maximal capacity of all connections between s_i and C via r_j is equal to the maximal capacity of all connections between C and s_i in the reverse direction, i.e. we look for x_j such that

$$\max_{j\in N}\{\min\{a_{ij}, x_j\}\} = \max_{j\in N}\{\min\{b_{ij}, x_j\}\} \text{ for all } i \in N.$$

For practical reasons, the reduction to a fixed control level λ is usually made and then the above formula has a matrix-vector form $A \otimes x = \lambda \otimes B \otimes x$ (a generalized eigenproblem), where $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

The eigenproblem in a max-min (fuzzy) algebra has been described in many monographs and papers, see [1], [2], [3], [10], [17].

[1] Technical University, Department of Mathematics and Theoretical Informatics, Nemcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

This paper generalizes the results presented in [12] and investigates the properties of generalized eigenspace for matrices and vectors with interval coefficients. An interval vector $\boldsymbol{X}$ is said to be a strong generalized eigenvector of square matrices $A,\ B$ if $A \otimes x = \lambda \otimes B \otimes x$ holds for each x in $\boldsymbol{X}$ and for some $\lambda \in \mathbb{B}$. We suppose that an interval matrix $\boldsymbol{A}$ can be split into two subsets according to forall–exists quantification of its interval entries. The properties of of generalized eigenvectors, namely EA-strong generalized eigenvectors are studied and characterizations of equivalent conditions are presented.

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 natural numbers (natural numbers with zero) is denoted by $\mathbb{N}$ ($\mathbb{N}_0$). For given $m, n \in \mathbb{N}$, write $M = \{1, 2, \ldots, m\}$ and $N = \{1, 2, \ldots, n\}$. The set of $m \times n$ matrices over $\mathbb{B}$ is denoted by $\mathbb{B}(m, n)$, and in particular, the set of $n \times 1$ vectors over $\mathbb{B}$ is denoted by $\mathbb{B}(n)$.

The operations $\oplus, \otimes$ are extended to the matrix-vector algebra over $\mathbb{B}$ by the direct analogy to the conventional linear algebra. If each entry of a matrix $A \in \mathbb{B}(n, n)$ (a vector $x \in \mathbb{B}(n)$) is equal to O we shall denote it as $A = O$ ($x = O$).

Let $x = (x_1, \ldots, x_n) \in \mathbb{B}(n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{B}(n)$ be vectors. We write $x \leq y$ if $x_i \leq y_i$ holds for each $i \in N$.

For a given matrix $A \in \mathbb{B}(n, n)$, the number $\lambda \in \mathbb{B}$ and the n-tuple $x \in \mathbb{B}(n)$ are the so-called *generalized eigenvalue* and *generalized eigenvector* of (A, B), respectively, if $A \otimes x = \lambda \otimes B \otimes x$.

An *eigenspace* $V(A, B, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e.,

$$V(A, B, \lambda) = \{x \in \mathbb{B}(n);\ A \otimes x = \lambda \otimes B \otimes x\}.$$

2 Interval Versions of Generalized Eigenvectors

Analogously to [2], [4]-[10], [12] consider interval matrices $\boldsymbol{A}$ with bounds $\underline{A}, \overline{A} \in \mathbb{B}(m, n)$ and an interval vector $\boldsymbol{X}$ with bounds $\underline{x}, \overline{x} \in \mathbb{B}(n)$ which are defined as follows:

Definition 1. Let $\underline{A} = (\underline{a}_{ij})$, $\overline{A} = (\overline{a}_{ij}) \in \mathbb{B}(n, n)$, $\underline{A} \leq \overline{A}$, $\underline{B} = (\underline{b}_{ij})$, $\overline{B} = (\overline{b}_{ij}) \in \mathbb{B}(n, n)$, $\underline{B} \leq \overline{B}$ and $\underline{x}, \overline{x} \in \mathbb{B}(n)$, $\underline{x} \leq \overline{x}$. An interval matrix $\boldsymbol{A}$ with bounds $\underline{A}, \overline{A}$, an interval matrix $\boldsymbol{B}$ with bounds $\underline{B}, \overline{B}$ and interval vector $\boldsymbol{X}$ are defined as follows: $\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathbb{B}(n, n);\ \underline{A} \leq A \leq \overline{A} \right\}$, $\boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \mathbb{B}(m, n);\ \underline{B} \leq B \leq \overline{B} \right\}$, $\boldsymbol{X} = [\underline{x}, \overline{x}] = \left\{ x \in \mathbb{B}(n);\ \underline{x} \leq x \leq \overline{x} \right\}$.

Suppose that each interval of $\boldsymbol{A}$ is associated either with the universal, or with the existential quantifier. Then we can split the interval matrix as $\boldsymbol{A} = \boldsymbol{A}^{\forall} \oplus \boldsymbol{A}^{\exists}$, where $\boldsymbol{A}^{\forall}$ is the interval matrix comprising universally quantified coefficients and $\boldsymbol{A}^{\exists}$ concerns existentially quantified coefficients. Thereafter denote by $\tilde{N}^{\exists} \subseteq N \times N$ and $\tilde{N}^{\forall} \subseteq N \times N$ the corresponding sets of indices. In other words, $\underline{a}_{ij}^{\exists} = \overline{a}_{ij}^{\exists} = 0$ for each couple $(i, j) \in \tilde{N}^{\forall}$ and $\underline{a}_{ij}^{\forall} = \overline{a}_{ij}^{\forall} = 0$ for each couple $(i, j) \in \tilde{N}^{\exists}$.

Example 1. Suppose that $\mathbb{B} = [0, 1]$. Consider interval vector $\boldsymbol{A}$. Then $\boldsymbol{A}^{\exists},\ \boldsymbol{A}^{\forall}$ have the forms

$$\boldsymbol{A} = \begin{pmatrix} [0.1, 0.2] & [0.1, 0.3] \\ [0.1, 0.2] & [\,0, 0.1] \\ [0.3, 0.4] & [0.1, 0.3] \end{pmatrix}\ \boldsymbol{A}^{\exists} = \begin{pmatrix} [0, 0] & [0.1, 0.3] \\ [0, 0] & [0, 0] \\ [0.3, 0.4] & [0, 0] \end{pmatrix} \text{ and } \boldsymbol{A}^{\forall} = \begin{pmatrix} [0.1, 0.2] & [0, 0] \\ [0.1, 0.2] & [\,0, 0.1] \\ [0, 0] & [0.1, 0.3] \end{pmatrix},$$

where $\tilde{N}^{\exists} = \{(1, 2), (3, 1)\}$ and $\tilde{N}^{\forall} = \{(1, 1), (2, 1), (2, 2), (3, 2)\}$.

Definition 2. If the interval matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and interval vector $\boldsymbol{X}$ are given, then $\boldsymbol{X}$ is called

- *a strong generalized eigenvector* of $(\boldsymbol{A}, \boldsymbol{B})$
  if $(\exists \lambda \in \mathbb{B})(\forall A \in \boldsymbol{A})(\forall B \in \boldsymbol{B})(\forall x \in \boldsymbol{X})[\, A \otimes x = \lambda \otimes B \otimes x\,]$,
- *an L-controllable generalized eigenvector* of $(\boldsymbol{A}, \boldsymbol{B})$
  if $(\exists \lambda \in \mathbb{B})(\exists A \in \boldsymbol{A})(\forall B \in \boldsymbol{B})(\forall x \in \boldsymbol{X})[\, A \otimes x = \lambda \otimes B \otimes x\,]$.

For given indices $i \in M, j \in N$ we define $\tilde{A}^{(ij)} \in \mathbb{B}(m, n)$, $\tilde{B}^{(ij)} \in \mathbb{B}(m, n)$ and $\tilde{x}^{(i)}$ by putting for every $k \in M, l \in N$

$$\tilde{a}_{kl}^{(ij)} = \begin{cases} \overline{a}_{ij}, & \text{for } k = i,\ l = j \\ \underline{a}_{kl}, & \text{otherwise} \end{cases}, \qquad \tilde{b}_{kl}^{(ij)} = \begin{cases} \overline{b}_{ij}, & \text{for } k = i,\ l = j \\ \underline{b}_{kl}, & \text{otherwise} \end{cases},$$

$$\tilde{x}_k^{(i)} = \begin{cases} \overline{x}_i, & \text{for } k = i \\ \underline{x}_k, & \text{otherwise} \end{cases} .$$

Lemma 1. [12] Let $x \in \mathbb{B}(n)$ and $A \in \mathbb{B}(m,n)$. Then

- $x \in \boldsymbol{X}$ if and only if $x = \bigoplus_{i \in N} \beta_i \otimes \tilde{x}^{(i)}$ for some values $\beta_i \in \mathbb{B}$ with $\underline{x}_i \le \beta_i \le \overline{x}_i$,
- $A \in \boldsymbol{A}$ if and only if $A = \bigoplus_{i \in M, j \in N} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ for some values $\alpha_{ij} \in \mathbb{B}$ with $\underline{a}_{ij} \le \alpha_{ij} \le \overline{a}_{ij}$.

Theorem 1. [12] Suppose given $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{X}$. Then $\boldsymbol{X}$ is a strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if there is $\lambda \in \mathbb{B}$ such that

$$\tilde{A}^{(ij)} \otimes \tilde{x}^{(t)} = \lambda \otimes \tilde{B}^{(rs)} \otimes \tilde{x}^{(t)} \quad \text{for every } i, r \in M, j, s, t \in N, \tag{1}$$

$$\bigoplus_{k \in M} \lambda \otimes (\overline{B} \otimes \overline{x})_k \le \bigoplus_{i \in M, j \in N} \underline{a}_{ij}, \tag{2}$$

$$\bigoplus_{k \in M} (\overline{A} \otimes \overline{x})_k \le \bigoplus_{i \in M, j \in N} \lambda \otimes \underline{b}_{ij}. \tag{3}$$

Theorem 2. [12] *Suppose given $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{X}$. Then $\boldsymbol{X}$ is an L-controllable generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if there exist $\lambda \in \mathbb{B}$ and $A \in \boldsymbol{A}$ such that $A \otimes \tilde{x}^{(k)} = \lambda \otimes \underline{B} \otimes \tilde{x}^{(k)}$ and $A \otimes \tilde{x}^{(k)} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(k)}$ for all $k \in N$.*

2.1 EA-Strong Generalized Eigenvectors

Definition 3. Let $\boldsymbol{A}^{\exists}$, $\boldsymbol{A}^{\forall}$, $\boldsymbol{B}$, $\boldsymbol{X}$ be given. Interval vector $\boldsymbol{X}$ is called *EA–strong generalized eigenvector* of $(\boldsymbol{A}, \boldsymbol{B})$ if there are $\lambda \in \mathbb{B}$ and $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that for any $A^{\forall} \in \boldsymbol{A}^{\forall}$, $B \in \boldsymbol{B}$ and $x \in \boldsymbol{X}$ the equality $(A^{\exists} \oplus A^{\forall}) \otimes x = \lambda \otimes B \otimes x$ holds.

Theorem 3. *Suppose given $\boldsymbol{A}^{\exists}$, $\boldsymbol{A}^{\forall}$, $\boldsymbol{B}$, $\boldsymbol{X}$. Then $\boldsymbol{X}$ is EA–strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if* $(\exists \lambda \in \mathbb{B})(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\forall A^{\forall} \in \boldsymbol{A}^{\forall})(\forall B \in \boldsymbol{B})(\forall k \in N)[(A^{\exists} \oplus A^{\forall}) \otimes \tilde{x}^{(k)} = \lambda \otimes B \otimes \tilde{x}^{(k)}]$.

Proof. Suppose that there are $\lambda \in \mathbb{B}$, $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that $(\forall A^{\forall} \in \boldsymbol{A}^{\forall})(\forall B \in \boldsymbol{B})(\forall k \in N)[(A^{\exists} \oplus A^{\forall}) \otimes \tilde{x}^{(k)} = \lambda \otimes B \otimes \tilde{x}^{(k)}]$ and $x \in \mathbb{B}(n)$ is an arbitrary vector in $\boldsymbol{X}$. Then in view of Lemma 1, we have $x = \bigoplus_{k \in N} \gamma_k \otimes \tilde{x}^{(k)}$ with $\underline{x}_k \le \gamma_k \le \overline{x}_k$. Therefore,

$$(A^{\exists} \oplus A^{\forall}) \otimes x = (A^{\exists} \oplus A^{\forall}) \otimes \bigoplus_{k \in N} \gamma_k \otimes \tilde{x}^{(k)} = \bigoplus_{k \in N} \gamma_k \otimes ((A^{\exists} \oplus A^{\forall}) \otimes \tilde{x}^{(k)}) =$$

$$\bigoplus_{k \in N} \gamma_k \otimes (\lambda \otimes B \otimes \tilde{x}^{(k)}) = \lambda \otimes B \otimes \bigoplus_{k \in N} \gamma_k \otimes \tilde{x}^{(k)} = \lambda \otimes B \otimes x.$$

The converse implication is trivial.

Theorem 4. *Suppose given $\boldsymbol{A}^{\exists}$, $\boldsymbol{A}^{\forall}$, $\boldsymbol{B}$ and vector $x \in \boldsymbol{X}$. Then $\boldsymbol{X}$ is EA–strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if there are $\lambda \in \mathbb{B}$ and $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that for any $A^{\forall} \in \boldsymbol{A}^{\forall}$ the following holds:*

$$(A^{\exists} \oplus A^{\forall}) \otimes x = \lambda \otimes \underline{B} \otimes x = \lambda \otimes \overline{B} \otimes x. \tag{4}$$

Proof. Suppose that there are $\lambda \in \mathbb{B}$, $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that $(\forall A^{\forall} \in \boldsymbol{A}^{\forall})(\forall B \in \boldsymbol{B})(\forall x \in \boldsymbol{X})[(A^{\exists} \oplus A^{\forall}) \otimes x = \lambda \otimes B \otimes x]$. Let $x \in \mathbb{B}(n)$ and $B \in \boldsymbol{B}$ be an arbitrary vector in $\boldsymbol{X}$ and an arbitrary matrix in $\boldsymbol{B}$, respectively. Then the assertion is the consequence of the next formula:

$$(A^{\exists} \oplus A^{\forall}) \otimes x = \lambda \otimes \underline{B} \otimes x \le \lambda \otimes B \otimes x \le \lambda \otimes \overline{B} \otimes x = (A^{\exists} \oplus A^{\forall}) \otimes x.$$

The converse implication is trivial.

Theorem 5. *Suppose given $\boldsymbol{A}^{\exists}$, $\boldsymbol{A}^{\forall}$, $\boldsymbol{B}$, vector $x \in \boldsymbol{X}$. Then $\boldsymbol{X}$ is EA–strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if there are $\lambda \in \mathbb{B}$ and $A^{\exists} \in \boldsymbol{A}^{\exists}$ such that*

$$(A^{\exists} \oplus \tilde{A}^{\forall(ij)}) \otimes x = \lambda \otimes \underline{B} \otimes x = \lambda \otimes \overline{B} \otimes x \text{ for every } (i,j) \in \tilde{N}_0^{\forall}. \tag{5}$$

Proof. Suppose that $x \in \boldsymbol{X}$ and $(\exists \lambda \in \mathbb{B})(\exists A^{\exists} \in \boldsymbol{A}^{\exists})(\forall A^{\forall} \in \boldsymbol{A}^{\forall})(\forall B \in \boldsymbol{B})[(A^{\exists} \oplus \tilde{A}^{\forall(ij)}) \otimes x = \lambda \otimes \underline{B} \otimes x = \lambda \otimes \overline{B} \otimes x]$. Then, trivially, (5) is true.

Suppose now that (5) holds true and $x \in \boldsymbol{X}$, $A \in \boldsymbol{A}$. Then in view of Lemma 1 we have $A = \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ with $\underline{a}_{ij} \leq \alpha_{ij} \leq \overline{a}_{ij}$. Therefore,

$$(A^{\exists} \oplus A^{\forall}) \otimes x = (A^{\exists} \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \alpha_{ij} \otimes \tilde{A}^{\forall(ij)}) \otimes x = (A^{\exists} \otimes x) \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \alpha_{ij} \otimes (\tilde{A}^{\forall(ij)} \otimes x) \geq$$

$$A^{\exists} \otimes x \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \underline{a}_{ij} \otimes (\tilde{A}^{\forall(ij)} \otimes x) = A^{\exists} \otimes x \oplus (\bigoplus_{(i,j) \in \tilde{N}^{\forall}} \underline{a}_{ij} \otimes (\tilde{A}^{\forall(ij)}) \otimes x = (A^{\exists} \oplus \underline{A}) \otimes x = \lambda \otimes \underline{B} \otimes x = \lambda \otimes \overline{B} \otimes x.$$

On the other hand we get

$$(A^{\exists} \oplus A^{\forall}) \otimes x = (A^{\exists} \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \alpha_{ij} \otimes \tilde{A}^{\forall(ij)}) \otimes x \leq (A^{\exists} \otimes x) \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \tilde{A}^{\forall(ij)} \otimes x =$$

$$\bigoplus_{(i,j) \in \tilde{N}^{\forall}} A^{\exists} \otimes x \oplus \bigoplus_{(i,j) \in \tilde{N}^{\forall}} \tilde{A}^{\forall(ij)} \otimes x = \lambda \otimes \underline{B} \otimes x = \lambda \otimes \overline{B} \otimes x.$$

□

Theorem 6. *Suppose given $\boldsymbol{A}^{\exists}$, $\boldsymbol{A}^{\forall}$, $\boldsymbol{B}$ and $\boldsymbol{X}$. Then $\boldsymbol{X}$ is a strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only if there is $\lambda \in \mathbb{B}$ such that*

$$(A^{\exists} \oplus \tilde{A}^{\forall(ij)}) \otimes \tilde{x}^{(k)} = \lambda \otimes \underline{B} \otimes \tilde{x}^{(k)} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(k)} \text{ for every } (i,j) \in \tilde{N}_0^{\forall} \text{ and for every } k \in N. \tag{6}$$

Proof. The equivalence follows from Theorem 3 and Theorem 5. □

Denote $\tilde{N}^{\forall} = \{(\alpha_1, \beta_1), \ldots, (\alpha_p, \beta_p)\}$ and $\tilde{N}^{\exists} = \{(\gamma_1, \delta_1), \ldots, (\gamma_r, \delta_r)\}$ with $p + r = n^2$. To recognize the existence of $A \in \boldsymbol{A}$ in Theorem 6, put $\tilde{C} \in \mathbb{B}(n.r)$ and $\tilde{D} \in \mathbb{B}(n.r, p+1)$ as follows:

$$\tilde{C} = \begin{pmatrix} \underline{B} \otimes \tilde{x}^{(1)} \\ \vdots \\ \underline{B} \otimes \tilde{x}^{(1)} \\ \underline{B} \otimes \tilde{x}^{(2)} \\ \vdots \\ \underline{B} \otimes \tilde{x}^{(2)} \\ \vdots \\ \underline{B} \otimes \tilde{x}^{(n)} \\ \vdots \\ \underline{B} \otimes \tilde{x}^{(n)} \end{pmatrix}, \quad \tilde{D} = \begin{pmatrix} \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(1)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\gamma_1,\delta_1)} \otimes \tilde{x}^{(1)} \\ \vdots & & & & \\ \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(n)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\gamma_1,\delta_1)} \otimes \tilde{x}^{(n)} \\ \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(1)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\gamma_2,\delta_2)} \otimes \tilde{x}^{(1)} \\ \vdots & & & & \\ \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(n)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\gamma_2,\delta_2)} \otimes \tilde{x}^{(n)} \\ \vdots & & & & \\ \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(1)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(1)} & \tilde{A}^{(\gamma_r,\delta_r)} \otimes \tilde{x}^{(1)} \\ \vdots & & & & \\ \tilde{A}^{(\alpha_1,\beta_1)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\alpha_2,\beta_2)} \otimes \tilde{x}^{(n)} & \ldots & \tilde{A}^{(\alpha_p,\beta_p)} \otimes \tilde{x}^{(n)} & \tilde{A}^{(\gamma_r,\delta_r)} \otimes \tilde{x}^{(n)} \end{pmatrix}. \tag{7}$$

Consider the following max-min linear system

$$\tilde{D} \otimes y = \lambda \otimes \tilde{C} \tag{8}$$

where the variable vector $y \in \mathbb{B}(p+1)$ consists of the variables $y_{ij} \in \mathbb{B}$.
Denote

$$\Lambda_{\max} = \bigotimes_{k,l \in N,\, (\underline{B} \otimes \tilde{x}^{(k)})_l < (\overline{B} \otimes \tilde{x}^{(k)})_l} (\underline{B} \otimes \tilde{x}^{(k)})_l$$

whereby $\min \emptyset = I$. Notice that the system $\lambda \otimes \underline{B} \otimes \tilde{x}^{(k)} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(k)}$ is solvable if and only if $\lambda \leq \Lambda_{\max}$.

Theorem 7. *Suppose given $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{X}$. Then $\boldsymbol{X}$ is an EA-strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ if and only there is $\lambda \in \mathbb{B}$, $\lambda_{\min} \leq \Lambda_{\max}$ such that the max-min linear system $\tilde{D} \otimes y = \lambda \otimes \tilde{C}$ has a solution y satisfying the condition $\underline{a}_{ij} \leq y_{ij} \leq \overline{a}_{ij}$, for every $(i,j) \in \tilde{N}^{\exists}$ and $I \leq y_{p+1,p+1} \leq I$.*

Proof. Suppose that y is a solution of the linear system $\tilde{D} \otimes y = \lambda \otimes \tilde{C}$ satisfying the condition $\underline{a}_{ij} \leq y_{ij} \leq \overline{a}_{ij}$, for every $i \in M, j \in N$. Then the matrix $A \in \mathbb{B}(m,n)$ defined as the max-min linear combination

$$A = \bigoplus_{i \in M, j \in N} y_{ij} \otimes \tilde{A}^{(ij)} \tag{9}$$

belongs to the interval matrix $[\underline{A}, \overline{A}]$ in view of Lemma 1.

Moreover, from the equality $\tilde{D} \otimes y = \lambda \otimes \tilde{C}$, we have the following block equations for every fixed $i \in N$

$$\bigoplus_{k \in M, l \in N} \left(\tilde{A}^{(kl)} \otimes \tilde{x}^{(i)} \right) \otimes y_{kl} = \lambda \otimes \underline{B} \otimes \tilde{x}^{(i)} \Leftrightarrow$$
$$\bigoplus_{k \in M, l \in N} \left(y_{kl} \otimes \tilde{A}^{(kl)} \right) \otimes \tilde{x}^{(i)} = \lambda \otimes \underline{B} \otimes \tilde{x}^{(i)} \Leftrightarrow A \otimes \tilde{x}^{(i)} = \lambda \otimes \underline{B} \otimes \tilde{x}^{(i)}$$

and

$$\bigoplus_{k \in M, l \in N} \left(\tilde{A}^{(kl)} \otimes \tilde{x}^{(i)} \right) \otimes y_{kl} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(i)} \Leftrightarrow$$
$$\bigoplus_{k \in M, l \in N} \left(y_{kl} \otimes \tilde{A}^{(kl)} \right) \otimes \tilde{x}^{(i)} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(i)} \Leftrightarrow A \otimes \tilde{x}^{(i)} = \lambda \otimes \overline{B} \otimes \tilde{x}^{(i)}.$$

Thus, in view of Theorem 2, $\boldsymbol{X}$ is an L-controllable generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$.

For the converse implication, assume that $\boldsymbol{X}$ is an L-controllable generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ i.e., that there exist $\lambda \in \mathbb{B}$ and $A \in \boldsymbol{A}$ such that for each $B \in [\underline{B}, \overline{B}])$ and each $x \in [\underline{x}, \overline{x}])$ the equality $A \otimes x = \lambda \otimes B \otimes x$ holds true. By Lemma 1, there exist coefficients $\alpha_{ij} \in \mathbb{B}$, $i \in M, j \in N$ such that $A = \bigoplus_{i \in M, j \in N} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ and $\underline{a}_{ij} \leq \alpha_{ij} \leq \overline{a}_{ij}$. It is easy to verify that $y \in \mathbb{B}(mn, 1)$, where $y_{ij} = \alpha_{ij}$ for every $i \in M, j \in N$, satisfies $\tilde{D} \otimes y = \lambda \otimes \tilde{C}$. □

We consider a max-min linear system

$$C \otimes y = \lambda \otimes b, \tag{10}$$
$$\underline{y} \leq y \leq \overline{y}, \tag{11}$$

We define $N^{\star} \subseteq N$, $r = r(C) \in \mathbb{B}(n)$ and $\lambda_{\min} = \lambda_{\min}(C, b, \underline{y}) \in \mathbb{B}$ by putting

$$N^{\star} = \left\{ j \in N;\ (\exists k \in N)\, c_{kj} > b_k \right\},\ \ r_j = \max_{k \in N} c_{kj} \quad \text{for every } j \in N,$$

$$\lambda_{\min} = \bigoplus_{j \in N^{\star}} \underline{y}_j \oplus \bigoplus_{j \in N \setminus N^{\star}} \left(r_j \otimes \underline{y}_j \right).$$

Lemma 2. [2] Let matrix $C \in \mathbb{B}(n,n)$ and vectors b, $\underline{y}$, $\overline{y} \in \mathbb{B}(n)$ be given. If $\lambda \in \mathbb{B}$, $y \in \mathbb{B}(n)$ are such that $C \otimes y = \lambda \otimes b$ and $\underline{y} \leq y \leq \overline{y}$, then $\lambda_{\min} \leq \lambda$.

Theorem 7 reduces the problem whether $\boldsymbol{X}$ is an EA-strongly generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$ to the solvability problem of the system $\tilde{D} \otimes y = \lambda \otimes \tilde{C}$ with $\underline{a}_{ij} \leq y_{ij} \leq \overline{a}_{ij}$ with $\lambda_{\min} \leq \Lambda_{\max}$.

Theorem 8. [2] *Suppose given $C \in \mathbb{B}(r,t)$, $b \in \mathbb{B}(r)$ and $\underline{y}, \overline{y} \in \mathbb{B}(t)$. The problem of recognizing the solvability of bounded parametric max-min linear system $C \otimes y = \lambda \otimes b$ with bounds $\underline{y} \leq y \leq \overline{y}$, for some value of parameter $\lambda \in \mathbb{B}$, can be solved in $O(rt)$ time.*

Algorithm: EA-Strong Generalized Eigenvector

Input. $\boldsymbol{A} = [\underline{A}, \overline{A}]$, $\boldsymbol{B} = [\underline{B}, \overline{B}]$, $\boldsymbol{X} = [\underline{x}, \overline{x}]$.
Output. 'yes' in variable sge if $\boldsymbol{X}$ is an EA-strong generalized eigenvector of $(\boldsymbol{A}, \boldsymbol{B})$; 'no' in sge otherwise.
begin

1 Compute $\tilde{C}$, $\tilde{D}$, $\lambda_{\min}(\tilde{C}, \tilde{D})$, $\Lambda_{\max}$;

2 If $\tilde{D} \otimes y = \lambda_{\min} \otimes \tilde{C}$ with $\underline{a}_{ij} \leq y_{ij} \leq \overline{a}_{ij}$ is solvable and $\lambda_{\min}(\tilde{C}, \tilde{D}) \leq \Lambda_{\max}$ then sge='yes';

3 sge='no';
end

Theorem 9. *Suppose given* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{X}$. *The* **Algorithm** EA-Strong Generalized Eigenvector *correctly recognizes whether* $\boldsymbol{X}$ *is a strong generalized eigenvector of* $(\boldsymbol{A}, \boldsymbol{B})$ *in* $O(m^2n^3)$ *arithmetic operations.*

Proof. The computation of $\tilde{C}$ needs $O(mn^2)$ time and the computation of $\tilde{D}$ requires to compute products $A^{(ij)} \otimes \tilde{x}^{(k)}$ for all $i \in M, j, k \in N$, while each of them needs $O(mn)$ time. Therefore, the computation of $\tilde{D}$ is done in $O(m^2n^3)$. The computation of a system $\tilde{D} \otimes y = \lambda_{\min} \otimes \tilde{C}$ needs $O(mn \cdot mn)$ time, where $\tilde{C} \in \mathbb{B}(2mn)$, $\tilde{F} \in \mathbb{B}(2mn, mn)$ and $\underline{a}_{ij} \leq y_{ij} \leq \overline{a}_{ij}$ (see [2]). The complexity of all steps of the algorithm is $O(mn^2) + O(m^2n^3) + O(m^2n^2) = O(m^2n^3)$ elementary operations. □

References

[1] Gavalec, M. (2004). *Periodicity in Extremal Algebra.* Gaudeamus, Hradec Králové.

[2] Gavalec, M., Plavka, J. & Ponce, D. (2019). Strong tolerance of interval eigenvectors in fuzzy algebra. *Fuzzy Sets and Systems* 369, 145–156.

[3] Gottwald, S. (2001). *Treatise on Many-Valued Logics; Studies in Logic and Computation*. Research Studies Press.

[4] Molnárová, M. (2018). Possible and universal robustness of special classes of matrices with inexact data. Mathematical methods in economics. *Proceedings of the 36th international conference* (pp. 348–353). Prague.

[5] Molnárová, M. (2019). Fuzzy interval Monge matrices with respect to robustness. Mathematical methods in economics. *Proceedings of the 37th international conference* (pp. 409–414). České Budějovice.

[6] Myšková, H. (2012). On an algorithm for testing T4 solvability of fuzzy interval systems *Kybernetika* 48(5), 924–938.

[7] Myšková, H. (2012). An iterative algorithm for testing solvability of max-min interval systems. *Kybernetika* 48(5), 879–889.

[8] Myšková, H. & Plavka, J. (2019). X^{AE} and X^{EA} robustness of max–min matrices, *Discrete Applied Mathematics* 267, 142–150.

[9] Myšková, H. & Plavka, J. (2014). The robustness of interval matrices in max-plus algebra. *Linear Algebra and its Applications* 445, 85–102.

[10] Myšková, H. & Plavka, J. (2022). *Max-plus steady states in discrete event dynamic systems with inexact data*. Discrete Event Dynamic Systems. https://doi.org/10.1007/s10626-022-00359-3.

[11] Myšková, H. & Plavka, J. (2020). AE and EA robustness of interval circulant matrices in max-min algebra, *Fuzzy Sets and Systems* 384, 91–104.

[12] Plavka, J. Gazda, M. (2021). Generalized eigenproblem of interval max-min (fuzzy) matrices. *Fuzzy Sets and Systems* 410, 27–44.

[13] Plavka, J. (2001). On Eigenproblem for Circulant Matrices in Max-Algebra. *Optimization* 50, 477–483.

[14] Sanchez, E. (1978). Resolution of eigen fuzzy sets equations, *Fuzzy Sets and Systems* 1, 69–74.

[15] Terano, T. & Tsukamoto, Y. (1977): Failure diagnosis by using fuzzy logic. *Proc. IEEE Conference on Decision Control* (pp, 1390–1395). New Orleans, LA.

[16] Zadeh, L. A. (1971). Toward a theory of fuzzy systems. *In: R. E. Kalman, N. De Claris, Eds., Aspects of Network and Systems Theory (Hold, Rinehart and Winston)* (pp. 209–245) New York.

[17] Zimmermann, K. (1976). *Extremální algebra*. Ekonomicko-matematická laboratoř EÚ ČSAV, Praha.

A Bargaining Theory Application in a Coordinated Closed Loop Supply Chain

Petr Pokorny[1]

Abstract. In this paper we present a Closed Loop Supply Chain (CLSC) model where a retailer sells products manufactured by a manufacturer. The products can be returned like returnable packaging (either disposable or reusable) used in the beverage industry. Both CLSCs can choose to work together through a revenue-sharing contract to improve their profits. The contract, once signed, sets limits on the revenue share within which the members can negotiate. It is assumed that the market with returns is growing, which provides a further scope for bargaining. We show that the share of revenue that the retailer can retain is smaller when both members want to cooperate, when the manufacturer is the leader, but the retailer is able to negotiate slightly better terms for its share of surplus profit as the returns market grows.

Keywords: Closed-Loop Supply Chain, Bargaining Theory, Game Theory, Nash Equilibrium, Revenue Sharing

JEL Classification: C72
AMS Classification: 91A06 91A10

1 Introduction

The concept of Closed-Loop Supply Chains (CLSC) has been gaining attention in supply chain management. CLSC considers not only the traditional forward flow of goods, but also the reverse flow of products at the end of their use/life. This paper focuses on the bargaining process when a manufacturer $\bar{M}$ and a retailer $\bar{R}$ decide to cooperate in order to increase their profits. We start with a Stackelberg manufacturer-led model, which yields a lower profit than an integrated CLSC. To improve the situation they decide to close a so-called revenue sharing (RS) contract. In the RS contract, the retailer shares a portion of its revenues with the manufacturer in exchange for a reduced wholesale price. This co-ordination mechanism can improve the manufacturer-led CLSC to produce a negotiable profit surplus. We will also consider the situation that the market for returned packaging can and should grow. This is addressed in the Directive (EU 2019/904), better known as the Single-Use Plastic Directive (SUPD), which sets a target of 90% separate collection of plastic bottles by 2029. In this study, we will focus on the whole bargaining process, starting with both members setting the minimum acceptable profits, concluding a revenue sharing contract with constraints that condition the creation of surplus profits. And finally, we will show how the two players bargain over the surplus in order to maximise their utility as well as that of the CLSC using the well-known Nash bargaining solution [10],[11].

1.1 Bargaining and Revenue Sharing Contract in CLSC

Due to space limitations, we present only the most relevant papers dealing with revenue sharing contracts and bargaining. Coordination in traditional SC or CLSC has been a key issue in recent years to improve competitiveness and find ways to share profits [6]. The classic works analysing the benefits and mechanisms of RS contracts are [1],[2]. In [7], the authors derive the equilibrium for a three-echelon SC with random demand. A structured overview of CLSC models including revenue sharing contracts is given in [5]. The authors show a positive contribution of an RS contract in both a two and a three echelon reverse chain with used computers in [8]. An interesting study on the administrative costs associated with RS contracts is presented in [4] to ensure rational management insights. In [3], the non-cooperative contracting mechanism is compared with the cooperative bargaining approach and they are applied to a two-product dual competing CLSC.

In [10], the non-cooperative contracting mechanism is compared with the cooperative bargaining approach and applied to a two-product, dual-competing CLSC. Recent work on bargaining in CLSC includes [12], which studies the bargaining process between a retailer and a third-party collector negotiating a collection fee. In [9], the authors apply the Nash bargaining fairness reference to the revenue sharing contract with respect to the bargaining power of the CLSC members. The authors compare the different bargaining solutions, namely the generalised Kalai-

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, nám. W. Churchilla 1938/4, 130 67 Praha 3, pokornyp@vse.cz.

Smorodinsky (KS), with a generalised Nash bargaining solution on a multi-level serial supply chain where the members negotiate the wholesale prices and show the dominance of the Nash solution in [13].

2 Model

Consider a CLSC consisting of a manufacturer ($\bar{M}$) and a retailer ($\bar{R}$). ($\bar{M}$) produces new products and sells them to $\bar{R}$ and also recycles the products at the end of their life (used packaging). $\bar{R}$ sells to the final consumer, purchases the used empty packaging and refunds the deposit to the consumer.

2.1 Notation and Assumptions

The CLSC parameters and decision variables are

c_n/c_r	Unit production costs of products made from new/recycled raw materials
c_e	Greening effort cost
q	Demand quantity for new products
r	Return quantity of used products
w	Wholesale price charged for the new products by $\bar{M}$ to $\bar{R}$
p	Retail price charged for the new products by $\bar{R}$ to customers
b	Buy-back price offered for the returns by $\bar{M}$ to $\bar{R}$
ε	Acquisition price (deposit) offered for the returns by $\bar{R}$ to customers
g	Greening investment (effort)
τ	EUC's minimum set return ratio on returnable packaging under deposit/refund scheme
N_p	Minimum net price for a customer as $N_p = p - \varepsilon$

π_j^i is a profit function of agent $j, j \in \{\bar{M}, \bar{R}, SC\}$ in model i, $i \in \{C, M, RS\}$, $\pi_{SC}^i = \pi_{\bar{M}}^i + \pi_{\bar{R}}^i$. Model C is the integrated or centralized case with one decision maker. Model C is the integrated or centralised case with a single decision maker. Model M is a decentralised wholesale contracting model with the manufacturer as the Stackelberg leader. Finally, we will coordinate the wholesale contract with the revenue sharing contract in the models called RS

2.2 CLSC Demand, Supply and Profit Functions

In CLSCs we distinguish demand and supply functions. The demand function capture the retail end-consumer/customer behavior and in our case can be formulated as $q^i = \alpha - \beta p^i + \gamma \varepsilon^i + \delta g^i$, where the parameters $\alpha, \beta, \gamma, \delta \geq 0$ are all positive, α is the total neutral market size for new material made products , β is the price sensitivity, γ is the customer's elasticity to the acquisition price ε, and δ captures the positive greening effort. The supply function is $r^i = \theta + \eta \varepsilon^i$, where θ is the neutral size of the market for returns and η is the price sensitivity of returns.

Centralized CLSC – Model C

The condition $q^i \geq r^i$ states that the products sold to customers must be higher than the number of market returns. Therefore, the SUPD condition for the minimum return rate of reusable packaging under the deposit and refund system can be set $\tau q^i \leq r^i$. In the centralized model is a single decision maker manages the entire CLSC. The profit function of the integrated CLSC can be written as

$$\max_{p^i, g^i, \varepsilon^i} \pi_{SC}^C(p^C, g^C, \varepsilon^C) = (p^C - c_n)q^C + (c_n - c_r - \varepsilon^C)r^C - c_e(g^C)^2, s.t. \quad \tau q^C \leq r^C \tag{1}$$

We have to formulate the Lagrangian function to solve this problem

$$\pi_L^C = (p^C - c_n)q^C + (c_n - c_r - \varepsilon^C)r^C - c_e g^2 - \lambda^C(r^C - \tau q^C), \tag{2}$$

where $\lambda \geq 0$ is the Lagrange multiplier. The maximization problem is now formulated as follows

$$\max_{p^C,\varepsilon^C,g^C} \pi_L^C \quad \text{s.t. } \lambda^C \geq 0,\ q^C \geq r^C, r^C - \tau q^C = 0 \tag{3}$$

Decentralized CLSC – Model M

If the manufacturer is the leader, then the optimization problem can be formulated as follows:

$$\max_{w^M,b^M,g^M} \pi_M^M(w^M, b^M, g^M) = (w^M - c_n)q^M + (c_n - c_r - b^M)r^M - c_e(g^M)^2, \tag{4}$$

$$s.t. \quad \tau q^M \leq r^M\ ,\ \max_{p^M,\varepsilon^M} \pi_R^M(p^M, \varepsilon^M) = (p^M - w^M)q^M + (b^M - \varepsilon^M)r^M. \tag{5}$$

As in the case of model C, we use the Lagrangian function and backward induction to obtain the equilibria of the decision variables.

Coordinated CLSC – Revenue Sharing Contract – Model RS

Finally, we present the coordination model that aims to increase the performance of the leader-follower model M to the level of model C. If the producer is the Stackleberg leader, then the application of the revenue sharing contract to coordinate the CLSC can be formulated as follows:

$$\max_{w^{RS},b^{RS},g^{RS}} \pi_M^{RS}(w^{RS}, b^{RS}, g^{RS}) = (w^{RS} - c_n)q^{RS} + (c_n - c_r - b^{RS})r^{RS} - c_e(g^{RS})^2 + (1 - t^{RS})p^{RS}q^{RS}, \tag{6}$$

$$s.t. \quad \max_{p^{RS},\varepsilon^{RS}} \pi_R^{RS}(p^{RS}, \varepsilon^{RS}) = p^{RS}q^{RS}t^{RS} - w^{RS}q^{RS} + (b^{RS} - \varepsilon^{RS})r^{RS}\ , \tag{7}$$

$$g^{RS} = g^{C*},\ p^{RS} = p^{C*}, \varepsilon^{RS} = \varepsilon^{C*}, \qquad 0 \leq t^{RS} \leq 1. \tag{8}$$

The parameter t^{RS}, which controls the share of revenue retained by /shared with $\bar{R}/\bar{M}$ is the subject to the negotiations.

2.3 Bargaining Process

The parameter t^{RS}, which controls the share of revenue retained by /shared with $\bar{R}/\bar{M}$ is the subject to the negotiation. In the bargaining process, both CLSC members act with individual rationality, so there are limits to $t^{RS*} \in < t_{LB}^{M*}, t_{UB}^{M*} >$ at which the RS contract can be signed, expressed as $\pi_{\bar{M}}^{RS*} \geq \pi_{\bar{M}}^{M*} \wedge \pi_{\bar{R}}^{RS*} \geq \pi_{\bar{R}}^{M*}$. Since the RS model improves the overall performance of the CLSC, $\pi_{SC}^{RS*} = \pi_{SC}^{C*}$, as if it were an integrated chain, there is a surplus profit that can be negotiated. Moreover, if $\bar{M}$ is the market leader then $t_{LB}^{M*}(\tau)$, $t_{UB}^{M*}(\tau)$ are decreasing functions of the return market size τ. This makes sense because $\bar{R}$ has a weaker position as the market matures and grows with the unique position of $\bar{M}$. Hence, $t^{RS*}(\tau) \in < t_{LB}^{M*}(\tau), t_{UB}^{M*}(\tau) >$, where $0 \leq \tau \leq 1$. The $t^{RS*}(\tau)$ is an important part of the bargaining process because it conditions the deal. The negotiation process can be summarised in the following steps:

1. Both CLSC members $\bar{R}$ and $\bar{M}$ know their minimum guaranteed profits $\pi_{\bar{R}}^{M*}$, $\pi_{\bar{M}}^{M*}$ based on the model M.
2. They enter the RS contract negotiations in order to obtain excess profits by $\bar{R}$ sharing $0 \leq t^{RS} \leq 1$ of their profits with $\bar{M}$ in exchange for the wholesale price reductions, and reach an agreement that the share should be limited $t^{RS*}(\tau) \in < t_{LB}^{M*}(\tau), t_{UB}^{M*}(\tau) >$ and depend on the yield market evolution captured in τ. In this case $t_{LB}^{M*}(\tau)$ corresponds to the most preferred t^{RS}level by $\bar{M}$, since in this case he receives the highest possible share of the retailer's revenues that $\bar{R}$ is willing to accept. At the same time, the RS contract is constrained by the upper bound t_{UB}^{M*}, which is preferred by $\bar{R}$ because it allows them to keep the highest possible share of their revenues that $\bar{M}$ is willing to accept.
3. Completion of the RS contract will improve the performance of the CLSC and generate a profit surplus to be negotiated between the partners. The allocation of the surplus will also depend on the relative bargaining power. To illustrate the bargaining process and the resulting new profits for $\bar{M}$ and $\bar{R}$, we will use the Nash bargaining solution.

2.4 Nash Bargaining Solution in a Revenue Sharing Contract

We will use the well-known Nash Bargaining Solution (NBS) applied to the CLSC revenue sharing solution. NBS states that the optimal outcome of a bargaining process is equal to maximising the surplus utility function $N = \max_{x,y}\big[(u(x) - u(g_x))\big(u(y) - u(g_y)\big)\big]$ where (g_x) , $u(g_y)$ are the utilities that players always receive if they choose not to bargain. When we apply NBS to RS contract, we set $u(x) = u(\bar{M}) = \phi\pi_{SC}^{RS*} - \pi_{\bar{R}}^{M*}$, $u(g_x) = 0$,

$u(y) = u(\bar{R}) = (1-\phi)\pi_{SC}^{RS*} - \pi_{\bar{M}}^{M*}$, $u(g_y) = 0$, $\phi \in \left\langle \frac{\pi_{\bar{R}}^{M*}}{\pi_{SC}^{RS*}}, \frac{\pi_{\bar{M}}^{M*}}{\pi_{SC}^{RS*}} \right\rangle = \langle \phi_{min}, \phi_{max} \rangle$. The utility function of $\bar{M}$ increases with the proportion of their profit that they negotiate out of the total CLSC profit, reduced by the guaranteed profit of $\bar{R}$ and vice versa. The optimization problem is then set as follows:

$$NS(\phi) = \max_{\phi} \left[\left(\phi \pi_{SC}^{RS*}(t^{RS*}(\tau)) - \pi_{\bar{R}}^{M*}(\tau) \right) \left((1-\phi)\pi_{SC}^{RS*}(t^{RS*}(\tau)) - \pi_{\bar{M}}^{M*}(\tau) \right) \right],$$

$$s.t. \quad \phi \in \langle \phi_{min}, \phi_{max} \rangle, t^{RS*}(\tau) \in < t_{LB}^{M*}(\tau), t_{UB}^{M*}(\tau) >. \tag{9}$$

Optimization problem (9) outputs a vector $\boldsymbol{\phi} = (\phi_1(\tau_1), \phi_2(\tau_2), \dots, \phi_n(\tau_n))$ of optimal parameters that reflect the different shares of excess profit received at different market return conditions. As the total profit of the CLSC is an increasing concave function of τ, so are the minimum profits of the players and the surplus profit of the RS-coordinated CLSC. Hence, the additional parameter τ captures the market dynamics and completes the bargaining process.

3 Numerical Example

Table 1 summarizes the parameters used in our example.

α	β	γ	δ	θ	η	c_r	c_n	c_e
220	0.6	0.2	2	20	6	10	30	200

Table 1 Model example parameters

3.1 Results Discussion

We begin by illustrating the outcome of the RS contract negotiations in Figure 1, which set limits on the share of revenue to be shared between $\bar{M}$ and $\bar{R}$. If both $\bar{M}$ and $\bar{R}$ agree, then the CLSC can be coordinated to achieve model C's profit, which exceeds the model M's total profit. And **Figure 2** shows the slightly improving bargaining outcome for retailer as the market with the returnable packaging grows.

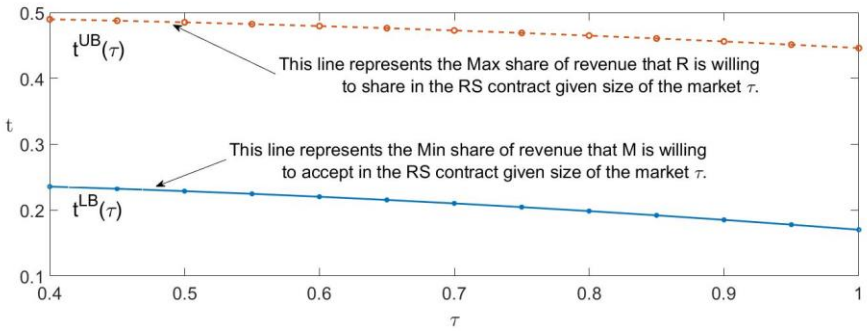


Figure 1 Results of the RS contract negotiations

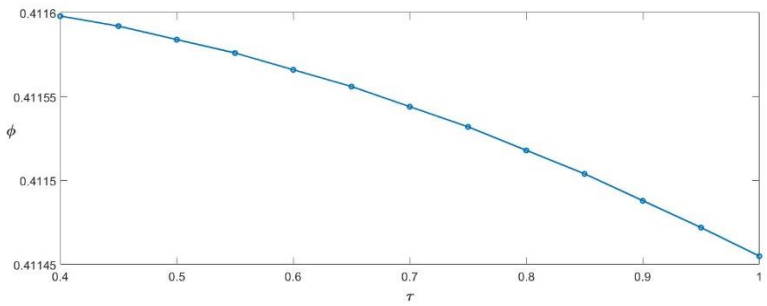


Figure 2 Optimal $\boldsymbol{\phi}$

Once the RS contract has been concluded, the players are free to negotiate the surplus gains achieved through coordination. Figure 3 shows in the left-hand panel the utility functions of $\bar{M}$ and $\bar{R}$, which move in opposite

directions according to (9). The utility of $\bar{M}$ increases as ϕ increases and that of $\bar{R}$ moves in the opposite direction. The two intersect around ϕ = 0.41, which maximises the total CLSC utility. As shown in **Figure 2**, ϕ evolves with the size of the market τ and decreases slightly, which has a negative impact on $\bar{M}$'s bargaining position.

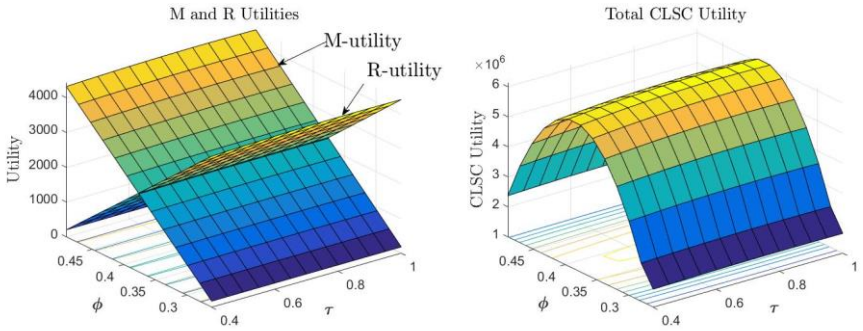


Figure 3 Utility Functions of $\bar{M}, \bar{R}$ and total CLSC

Finally, we present in Figure 4 the outcome of the bargaining process by comparing the old and new profits of both players. We can see that the manufacturer has managed to negotiate a higher share of the surplus profit $\phi \in \left\langle \frac{\pi_{\bar{R}}^{M*}}{\pi_{SC}^{RS*}}, \frac{\pi_{\bar{M}}^{M*}}{\pi_{SC}^{RS*}} \right\rangle = \langle \phi_{min}, \phi_{max} \rangle = \langle 0.25, 0.49 \rangle$, with $\phi^*(\tau) \cong 0{,}41$. This is a logical outcome of the model, the coordinated RS model is still a Stackelberg manufacturer-led CLSC where $\bar{M}$ knows the best response of $\bar{R}$ and thus controls the market.

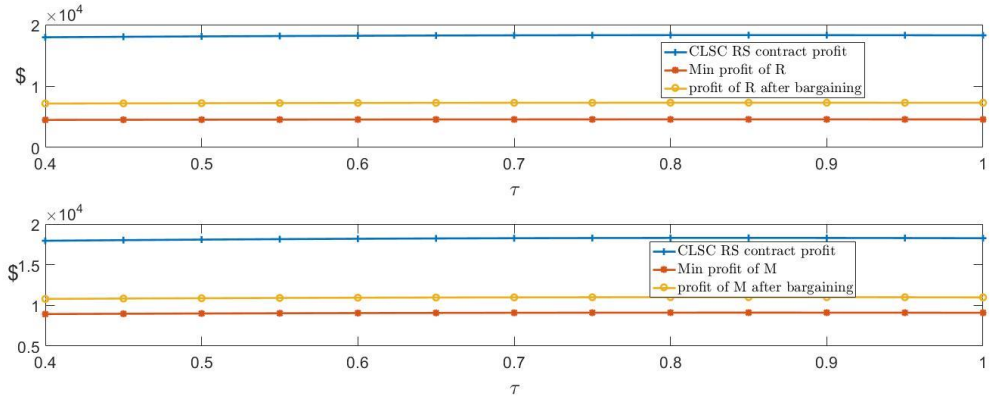


Figure 4 Profit functions after bargaining

4 Conclusion

In this paper we have focused on the bargaining process as a whole. The literature on CLSC coordination mainly focuses on the coordination mechanisms that lead to surplus profits. How these profits are allocated is what we have analysed in this paper. We followed the whole process, starting with both the manufacturer and the retailer agreeing to enter into the coordination deal, a revenue sharing contract, in order to improve performance and profits suffering from decentralisation and double marginalisation. The RS contract proved to be successful and generated a surplus profit for the CLSC as a whole, while still imposing restrictions on the minimum and maximum levels of revenue shared between members. In order to share the surplus, both entered into a bargaining game. The outcome is such that both the manufacturer and the retailer share the surplus profit and jointly maximise the system-wide CLSC utility. We have also shown that bargaining power does not change too much as the market grows, and still improves slightly in favour of the retailer.

Acknowledgements

The research project was supported by Grant No. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

[1] Cachon, G. (2003). Supply Chain Coordination with Contracts. *Handbooks in Operations Research and Management Science*, 11. https://doi.org/10.1016/S0927-0507(03)11006-7.

[2] Cachon, G. P. & Lariviere, M. A. (2005). Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science*, 51(1), pp. 30–44. https://doi.org/10.1287/mnsc.1040.0215.

[3] Chen, J. M. (2014). Coordinating a closed-loop supply chain using a bargaining power approach. *International Journal of Systems Science*, 40, pp. 1051–1061. https://doi.org/10.1080/23302674.2014.915356.

[4] De Giovanni, P. (2011). Environmental collaboration in a closed-loop supply chain with a reverse revenue sharing contract. *Annals of Operations Research - Annals OR*, 220. https://doi.org/10.1007/s10479-011-0912-5.

[5] De Giovanni, P. & Zaccour, G. (2019). A Selective Survey of Game-Theoretic Models of Closed-Loop Supply Chains. *4OR: A Quarterly Journal of Operations Research*, 17(1), pp. 1–44.

[6] Fiala, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research: CEJOR*, 24(2), pp. 267–281. https://10.1007/s10100-015-0423-6.

[7] Giannoccaro, I. & Pontrandolfo, P. (2004). Supply chain coordination by revenue sharing contracts. *International Journal of Production Economics*, 89(2), pp. 131–139. https://doi.org/10.1016/S0925-5273(03)00047-1.

[8] Govindan, K. & Popiuc, M. N. (2014). Reverse supply chain coordination by revenue sharing contract: A case for the personal computers industry. *European Journal of Operational Research*, 233(2), pp. 326–336.

[9] Li, Z. et al. (2022). Coordination of a supply chain with Nash bargaining fairness concerns. *Transportation Research Part E Logistics and Transportation Review*, 159. https://doi.org/10.1016/j.tre.2022.102627.

[10] Nash, J. (1953). Two-Person Cooperative Games. *Econometrica*, 21(1), pp. 128–140. https://doi.org/10.2307/1906951.

[11] Nash, J. F. (1950). The Bargaining Problem. *Econometrica*, 18(2), pp. 155–162. https://doi.org/10.2307/1907266.

[12] Tanai, Y. et al. (2019). Bargaining in a closed-loop supply chain with consumer returns. *International Journal of Operational Research*. https://doi.org/10.1504/IJOR.2021.115624.

[13] Zhong, F., Xie, J. & Shen, Y. (2022). Bargaining in a multi-echelon supply chain with power structure: KS solution vs. Nash solution. *Journal of Industrial and Management Optimization*. https://doi.org/10.3934/jimo.2020172.

Calculating Desirable Properties In MCDM

Jaroslav Ramík[1]

Abstract. Pairwise comparisons matrices (PCMs) are inevitable tools in some important multiple criteria decision making methods, e.g. AHP, TOPSIS, PROMETHEE and others. In this paper we investigate some important properties of PCMs which influence the generated priority vectors for final ranking the given alternatives. The novelty of our approach is that the priority vector is calculated as the solution of an optimization problem where an error objective function is minimized subject to constraints given by the desirable properties. The properties of the optimal solution are discussed and some illustrating examples are presented.

Keywords: multi-criteria decision making (MCDM), pairwise comparison matrix, consistency, coherence, priority vector

JEL Classification: C44, C63
AMS Classification: 90C29, 90C70

1 Introduction

A fundamental problem of decision theory is how to derive the weights for a set of objects (alternatives, activities, etc.) according to their importance, which are usually judged in pairs according to several criteria. By these weights, the objects are then ordered. This is a process of multiple criteria decision making (MCDM) which is a theory of measurement in a hierarchical structure consisting of the goal, the criteria (subcriteria), and the alternatives, see [6].

Pairwise comparisons matrices (PCMs) are important tools in many well known multiple criteria decision making methods, e.g. AHP, TOPSIS, PROMETHEE, etc., see [5]. In this paper we investigate some important and natural properties of PCMs called the *desirable properties*, particularly, the non-dominance, consistency, intensity and coherence, which influence the generated priority vectors for final ranking of the given alternatives.
The novelty of our approach is that the priority vector is calculated as the solution of an optimization problem where a special error objective function is minimized subject to constraints given by the desirable properties. The purpose is to derive the priority vector characterizing the proper ranking of elements.
Comparing to [6], and [5], here, we propose newly reformulated desirable properties – the non-dominance, consistency, intensity, and coherence – of the priority vector and we also propose an optimization method how to generate priority vectors with these desirable properties of the given pairwise comparison matrix.

2 Preliminaries

The reader can find the corresponding basic definitions, concepts and results, e.g. in [4]. Here, we summarize some necessary concepts. For detailed information, we refer to [5].
Given a set $\mathscr{C} = \{c_1, c_2, \ldots, c_n\}$ of elements, and corresponding $n > 1$, and an $n \times n$ pairwise comparison matrix, $A = \{a_{ij}\}$, whose entries, a_{ij}, evaluate the relative importance of the elements with respect to a given criterion.
Let $\mathscr{G} = (G, \cdot, \leq)$ be a multiplicative alo-group over an open interval, $G =]0; +\infty[$ of the set of real numbers $\mathbf{R}$. Here, the division operation "$\div$" on G is an inverse operation to the multiplicative operation "$\cdot$", for more details see [5]. Let $A = \{a_{ij}\}$ be an $n \times n$ matrix where each element a_{ij} belongs to G.

The matrix $A = \{a_{ij}\}$ is said to be *reciprocal* if the following condition holds for each $i, j \in \{1, \ldots, n\} = \mathscr{N}$: $a_{ij} \cdot a_{ji} = 1$.
If $A = \{a_{ij}\}$ is reciprocal, then $A = \{a_{ij}\}$ is called a *pairwise comparison matrix*, *PC matrix*, or, shortly, *PCM*.

A PC matrix $A = \{a_{ij}\}$ is said to be *consistent* if the following condition holds for each $i, j, k \in \mathscr{N}$: $a_{ik} = a_{ij} \cdot a_{jk}$.
The next proposition gives an equivalent condition for a PC matrix to be consistent, see, e.g. [5]. The proof of the following proposition is easy, or can be found in [5].

[1] Silesian University in Opava, SBA in Karviná, Dept. of Inf. and Math., University Sq. 1934/3, 73340 Karviná, Czechia, ramik@opf.slu.cz

Proposition 1. *Let $A = \{a_{ij}\}$ be a PC matrix. Then A is consistent if and only if (shortly: iff) there exists a vector $w = (w_1, \ldots, w_n)$ with $w_i \in G$, for $i \in \mathcal{N}$, such that for each $i, j \in \mathcal{N}$, it holds:*

$$a_{ij} \leq \frac{w_i}{w_j}. \tag{1}$$

The result of the pairwise comparisons method based on the PC matrix $A = \{a_{ij}\}$ is a rating of the set $\mathscr{C} = \{c_1, c_2, \ldots, c_n\}$ of the elements, i.e. a mapping that assigns real values to the elements (criteria or alternatives). Formally, it can be introduced as follows:
The *ranking function* for $\mathscr{C}$ (or the *ranking* of $\mathscr{C}$) is a function $w\colon \mathscr{C} \to G$ that assigns to every element from $\mathscr{C} = \{c_1, c_2, \ldots, c_n\}$ a value from the linearly ordered set G of the alo-group $\mathscr{G} = (G, \cdot, \leq)$.
Here, $w(c)$ represents the ranking value for $c \in \mathscr{C}$. The function w is usually written in the form of a vector of *weights*, i.e. $w = \big(w(c_1), w(c_2), \ldots, w(c_n)\big)$, or, simply $w = \big(w_1, w_2, \ldots, w_n\big)$, and it is called the *priority vector*. Also, we say that the priority vector w is associated with the PC matrix A, or that the priority vector w is generated by a priority generating method based on the PC matrix A.

The priority vector $w = \big(w_1, w_2, \ldots, w_n\big) \in G^n$ is *multiplicatively normalized*, if $\prod_{i=1}^{n} w_i = 1$.
Notice that if $w \in G^n$ is multiplicatively normalized then: $u = \big(\frac{w_1}{\sum_{i=1}^{n} w_i}, \ldots, \frac{w_n}{\sum_{i=1}^{n} w_i}\big)$ is additively normalized.

3 Priority Vectors of PC Matrices and Their Properties

Let $A = \{a_{ij}\}$ be a PC matrix on the alo-group $\mathscr{G} = (G, \cdot, \leq)$, let $w = (w_1, w_2, \ldots, w_n)$, with $w_j \in G$, be a priority vector.

(i) We say that the vector w is an *consistent vector* (CsV) of the PC matrix A if the following condition holds:

$$a_{ij} \leq \frac{w_i}{w_j} \qquad \text{for all} \quad i, j \in \mathcal{N}. \tag{2}$$

The PC matrix A is called a *consistent PC matrix*, if there exists a consistent vector w of the PC matrix A. Condition (2) is called the *consistent condition* (CsC).

(ii) We say that the vector w is an *intensity vector* (InV) of the PC matrix A if the following condition holds:

$$a_{ij} > a_{kl} \quad \text{iff} \quad \frac{w_i}{w_j} > \frac{w_k}{w_l} \qquad \text{for all} \quad i, j, k, l \in \mathcal{N}. \tag{3}$$

If there exists an intensity vector of the FPC matrix A, then A is called an *intensity FPC matrix*. Condition (3) is called the *intensity condition* (InC).

(iii) We say that the vector w is an *coherent vector* (CoV) of the PC matrix A if the following condition holds:

$$a_{ij} > 1 \quad \text{iff} \quad w_i > w_j \qquad \text{for all} \quad i, j \in \mathcal{N}. \tag{4}$$

If there exists a coherent vector of the PC matrix A, then A is called a *coherent PC matrix*. Condition (4) is called the *coherent condition* (CoC).

Remark 1. Notice that by the reciprocity property of the elements it is easy to see that w is a CsV of $A = \{\tilde{a}_{ij}\}$ if and only if

$$a_{ij} = \frac{w_i}{w_j} \qquad \text{for all} \quad i, j \in \mathcal{N}. \tag{5}$$

As will be shown in the following text of this section, some inconsistent pairwise comparisons matrices violate the non-dominant (ND) condition: the 'best' alternative is selected from the set of non-dominated alternatives, while this set is nonempty. Inconsistent PCMs that violate this natural condition should be viewed as logically flawed and should not be used for the derivation of weights of alternatives or other objects in MCDM. Other PCMs may violate the coherent condition (CoC), or the intensity condition (InC), see [1]. Furthermore, a new non-linear optimization problem is proposed for generating a priority vector (weights of alternatives, criteria, or other alternatives).

Let $A = \{a_{ij}\}$ be the PC matrix based on the set of alternatives $\mathscr{C} = \{c_1, c_2, \cdots, c_n\}$. We say that an alternative c_i *dominates alternative* c_j, and we write $c_i \succeq c_j$, or, equivalently, that an alternative c_j *is dominated by alternative* c_i, if $a_{ij} > 1$. 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 A is denoted by $ND(A) = \{c_j \in \mathscr{C} |$ there is no $i \in \mathcal{N} : c_i \succeq c_j\}$.

Let $A = \{a_{ij}\}$ be the PC matrix based on the set of alternatives $\mathscr{C} = \{c_1, c_2, \cdots, c_n\}$. Assume that $ND(A)$ is non-empty. Let $w = (w(c_1), ..., w(c_n))$ be the priority vector (i.e. vector of weights) associated to A. We say that the *non-dominated condition (NDC) is satisfied with respect to A and w*, if the maximal weight of the priority vector is associated with a non-dominated alternative.
Equivalently, we say that w satisfies the NDC with respect to A, if for some $i^* \in \{1, \cdots, n\}$:

$$c_{i^*} \in ND(A) \text{ and } w(c_{i^*}) = \max\{w(c_j) | j \in \{1, \cdots, n\}\}. \tag{6}$$

Alternatively, we say that A satisfies NDC with respect to w. The following example demonstrates a 'sufficiently' consistent PCM where Saaty's consistency index is less than 0.1, but the NDC and CoC conditions are, however, not met, see [4].

Example 1. Consider the set of four alternatives $\mathscr{C} = \{c_1, c_2, c_3, c_4\}$, and the corresponding PC matrix A given as follows:

$$A = \begin{pmatrix} 1 & 1.5 & 2 & 2 \\ 0.67 & 1 & 4 & 4 \\ 0.5 & 0.25 & 1 & 1 \\ 0.5 & 0.25 & 1 & 1 \end{pmatrix},$$

From the first row of PC matrix A, alternative c_1 clearly dominates the other three alternatives. Hence, c_1 is non-dominated. Saaty's consistency index CI and consistency ratio CR are: $CI = 0.052, RI = 0.089, CR = CI/RI = 0.058$. According to Saaty, see [6], for a 4×4 PCM, here, inconsistency is acceptable if $CR < 0.08$, (for an $n \times n$ PCM, $n > 4 : CR < 0.1$). Hence, inconsistency of A is acceptable and the priority vector (additively normalized) is generated by EVM as follows: $w_{EV} = (0.350, 0.396, 0.127, 0.127)$.
The weights of all alternatives (the priority vector w additively normalized) derived by GMM are as follows: $w_{GM} = (0.344, 0.396, 0.130, 0.130)$.
According to both EVM and GMM, the alternative with the highest weight is alternative c_2. This alternative is, however, dominated by alternative c_1, which is the only non-dominated alternative. Therefore, both w_{EV} and also w_{GM} violate NDC with respect to A, even though the consistency ratio $CR = 0.058$ is bellow Saaty's threshold of 0.08. Moreover, the Co condition is violated, too, as $a_{12} = \frac{3}{2} > 1$ and $\frac{w_1}{w_2} = \frac{0.344}{0.396} = 0.866 < 1$.

The following proposition says that the CoC is stronger than the NDC, i.e. CoC condition implies NDC. The opposite assertion does not hold as it is demonstrated in Example 1. The proof is evident.

Proposition 2. *Let $A = \{a_{ij}\}$ be a PC matrix, and let $w = (w_1, \cdots, w_n)$ be a PV associated to A. If A satisfies the CoC with respect to w, then A satisfies the NDC with respect to w.*

Remark 2. From definition it is clear that the In condition is stronger than the Co condition. In other words, if the In condition is satisfied, then the Co condition is satisfied, too. A PC matrix A from Example 1 violates the Co condition with respect to priority vector w generated by the GM method.

Remark 3. Let $A = \{a_{ij}\}$ be a consistent pairwise comparison matrix, and let $w = (w_1, \cdots, w_n)$ be a priority vector associated with A satisfying (5). Then it is obvious that ND, Co and In conditions are satisfied. Moreover, for a consistent pairwise comparison matrix, it is clear that the priority vector satisfying (5) can be generated by either EVM or by GMM.

Example 2. Consider a PC matrix D given as follows:

$$D = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 3 & 2 \\ 0.5 & 0.33 & 1 & 2 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix},$$

and the priority vector (obtained by the GM method, multiplicatively normalized): $w^{(D)} = (1.414, 1.565, 0.760, 0.595)$.
We obtain that $w_2 = 1.565$ is the maximal weight and c_2 is non-dominated alternative, hence the ND condition is met. Moreover, inconsistency of D is acceptable as $CR = 0.044 < 0.08$.
As it can be easily demonstrated, Co condition is satisfied:

$$d_{13} = 2, \frac{w_1}{w_3} = \frac{1.414}{0.760} > 1.000, d_{14} = 2, \frac{w_1}{w_4} = \frac{1.414}{0.595} > 1.000,$$

$$d_{23}=3,\frac{w_2}{w_3}=\frac{1.565}{0.760}>1.000,d_{24}=2,\frac{w_2}{w_4}=\frac{1.565}{0.595}>1.000,d_{34}=2,\frac{w_3}{w_4}=\frac{0.760}{0.595}>1.000.$$

On the other hand, the In condition is not met, e.g.:

$$d_{23}=3>d_{14}=2,\text{ and }\frac{w_2}{w_3}=\frac{1.565}{0.760}=2.060<\frac{w_1}{w_4}=\frac{1.414}{0.595}=2.378.$$

It is highly desirable that for a given PC matrix, A, possibly inconsistent, we are able to generate a priority vector w such that the ND, Co and In conditions are satisfied. For this purpose we shall formulate a special optimization problem whose solution will generate the desirable priority vector associated with the PC matrix A satisfying all above stated conditions, see also [3].

4 Deriving Priority Vectors of PC Matrices with the Desirable Properties

4.1 (Problem 0)

It was shown in Section 3, Example 1, that the calculation of a priority vector by the EV or GM methods from an inconsistent pairwise comparison matrix may result in violating the desirable conditions NDC, CoC, or InC. Therefore, an alternative approach to the derivation of a priority vector for PCMs may be formulated in terms of satisfying the ND, In and Co conditions.
Let $A=\{a_{ij}\}$ be a PC matrix. Based on this PCM, we need the following two sets of indexes:

$$I^{(2)}(A)=\{(i,j)|i,j\in\{1,\ldots,n\},a_{ij}>1\},\tag{7}$$

$$I^{(4)}(A)=\{(i,j,k,l)|i,j,k,l\in\{1,\ldots,n\},a_{ij}>1,a_{kl}>1,a_{ij}>a_{kl}\}.\tag{8}$$

Let $\delta:(x,y)\in\mathbf{R}_+\times\mathbf{R}_+\longrightarrow\delta(x,y)\in\mathbf{R}_+$ be a *distance function*, i.e. a function with the following well known properties for all $x,y,z\in\mathbf{R}_+$:

(i) $\delta(x,y)\geq 0$, , (ii) $\delta(x,y)=0$ iff $x=y$, (iii) $\delta(x,y)=\delta(y,x)$, (iv) $\delta(x,z)\leq\delta(x,y)+\delta(y,z)$.

Let $w=(w_1,\cdots,w_n)$ be a priority vector associated with A. An $n\times n$ *matrix of distances*, $\Delta(A,w)$, is defined as:

$$\Delta(A,w)=\{\Delta_{ij}\}=\{\delta(a_{ij},\frac{w_i}{w_j})\},$$

and a *matrix aggregation function*, $\Phi:X\in\mathbf{R}_+^n\times\mathbf{R}_+^n\longrightarrow\Phi(X)\in\mathbf{R}_+$, as an idempotent and increasing function (in each variable), where $X=\{x_{ij}\}$ is an $n\times n$ PC matrix.
An *error function*, $\mathscr{E}_A$, of $w=(w_1,\cdots,w_n)$ is defined as follows: $\mathscr{E}_A:w\in\mathbf{R}_+^n\longrightarrow\mathscr{E}_A(w)\in\mathbf{R}_+$,

$$\mathscr{E}_A(w)=\Phi(\Delta(A,w)).\tag{9}$$

The problem of finding a priority vector satisfying the ND, Co and In conditions can be formulated in terms of the following optimization problem, where $A=\{a_{ij}\}$ is a given PC matrix and $w=(w_1,\cdots,w_n)$ is an unknown priority vector with variables $w_1,\cdots,w_n\in G$:
(**Problem 0**)

$$\mathscr{E}_A(w)\longrightarrow\min;\tag{10}$$

subject to

$$\prod_{r=1}^{n}w_r=1,w_r>0\ \ r\in\mathscr{N},\tag{11}$$

$$w_r>w_s,\quad(r,s)\in I^{(2)}(A),\tag{12}$$

$$\frac{w_r}{w_s}>\frac{w_t}{w_u},\ \ (r,s,t,u)\in I^{(4)}(A).\tag{13}$$

The objective function in (10) minimizes the distance between the elements of PC matrix A and corresponding elements of the PCM $W=\{\frac{w_i}{w_j}\}$, measured by distance function δ. By constraint (11), the weights are positive and (multiplicatively) normalized. By (12), the Co condition is secured and by (13) the In condition is satisfied.

4.2 Transformation to (Problem ε)

Unfortunately, (Problem 0) has not been formulated in the form of a *standard optimization problem* that is appropriate for solving by existing numerical methods, see e.g. [2]. Here, variables w_i are required to be strictly positive and some inequality constraints, (12), (13), are strict, hence the set of feasible solution is not closed, as it is usual. That is why we transform the problem into a more convenient form by an appropriately small positive constant. Given a sufficiently small $\varepsilon > 0$.
(**Problem** ε)

$$\mathscr{E}_A(w) \longrightarrow \min; \tag{14}$$

subject to

$$\sum_{r=1}^{n} w_r = 1, w_r \geq \varepsilon, \;\; r \in \mathscr{N}, \tag{15}$$

$$w_r - w_s \geq \varepsilon, \;\; (r,s) \in I^{(2)}(A), \tag{16}$$

$$\frac{w_r}{w_s} - \frac{w_t}{w_u} \geq \varepsilon, \;\; (r,s,t,u) \in I^{(4)}(A). \tag{17}$$

In (Problem ε), nonlinear constraint (11) (with the product normalization) is substituted by a linear constraint (15) (with the additive normalization). Such a transformation is possible as the multiplicative and additive normalization formulas of priority vectors are equivalent. Notice that here, strict inequalities have been changed to the non-strict ones by adding a sufficiently small constant $\varepsilon > 0$. The proof of the next proposition is evident.

Proposition 3. *(Problem 0) has a feasible solution $w \in G^n$ if and only if there exists $\varepsilon > 0$ such that w is a feasible solution of (Problem ε).*
Moreover, if (Problem 0) has an optimal solution w^* then there exists $\varepsilon > 0$ such that w^* is an optimal solution of (Problem ε).

Here, by (Problem ε) we denote the following three optimization problems depending on the particular formulation of the objective function (14) as well as constraints (15) - (17), i.e. nested sets of feasible solutions. Some examples are presented bellow. We shall consider the following optimization problem variants:

(I) Minimize the objective function (14), subject to (15). The optimal solution is denoted by $w^{(I)}$. The ND, Co and In conditions are not necessarily satisfied.
(II) Minimize the objective function (14) subject to constraints (15), (16). The optimal solution is denoted by $w^{(II)}$. The Co condition is satisfied; then by Proposition 3 the ND condition is also satisfied. The In condition is not necessarily satisfied.
(III) Minimize the objective function (14) subject to constraints (15), (16), and (17). The optimal solution is denoted by $w^{(III)}$. Here, the ND, Co and In conditions should be satisfied.

4.3 Solving (Problem ε)

Notice that the set of feasible solutions of (Problem ε), (14) - (17), could be empty, e.g. for problems (II), and/or (III), see bellow. Even for a nonempty set of feasible solutions of (Problem 0), the optimal solution of the corresponding optimization problems (I), (II), or (III) need 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 $w^* = (w_1^*, \cdots, w_n^*)$ of some problems of (I) - (III) of (Problem ε) exists, the ND, Co, and In conditions hold by the nested properties of the feasible solution sets. Then, $w^* = (w_1^*, \cdots, w_n^*)$ is an appropriate priority vector associated with A satisfying the required properties. The proof is easy.

Proposition 4. *Let $A = \{a_{ij}\}$ be a* consistent *pairwise comparison matrix. Then there is a unique optimal solution $w^* = (w_1^*, \cdots, w_n^*)$ of (Problem 0) satisfying: $a_{ij} = \frac{w_i^*}{w_j^*}$ for all $i, j \in \mathscr{N}$ such that the ND, Co, and In conditions are met.*

Examples of distance functions $\delta(x,y)$: Let $x, y \in \mathbf{R}_+$.
(i) $\delta(x,y) = |x-y|$, (ii) $\delta(x,y) = (x-y)^2$, (iii) $\delta(x,y) = \max\{\frac{x}{y}, \frac{y}{x}\}$.
Examples of aggregation functions $\Phi(X)$: Let $X = \{x_{ij}\}$ be a $n \times n$ matrix, $x_{ij} \in \mathbf{R}_+$.
(a) $\Phi(X) = \frac{1}{n^2}\sum_{i,j=1}^{n} x_{ij}$, (b) $\Phi(X) = \max\{x_{ij} | i, j \in \mathscr{N}\}$.
Then the objective function in (Problem 0) and (Problem ε) is defined by (iii) and (b) as:

$$\mathscr{E}_A(w) = \max\{\max\{\frac{a_{ij}w_j}{w_i}, \frac{w_i}{a_{ij}w_j}\} | i, j \in \mathscr{N}\}. \tag{18}$$

For other combinations of functions δ and Φ our approach would need some modifications. When solving a particular optimization problem, (Problem ε), (14) - (17) with the objective function (18), we can encounter numerical difficulties, as this optimization problem is non-linear and also non-convex. Non-convexity is found in objective function (18) and also in constraints (17). Fortunately, these obstacles can be avoided by a proper approach - transformation of the non-convex problem to a convex one, which enables us using standard numerical methods for solving NLP problems. Then, for variants (I) and (II) of (Problem ε), we obtain an optimization problem solvable e.g by efficient interior point methods (see e.g. [2]). For solving variant (III) with non-convex constraints (17), we can apply e.g. an interior or exterior penalty method by penalizing this constraint and moving it into the objective function (see e.g. [2].

First, we analyze the objective function (18). Here, setting $f_{ij}(w) = \frac{a_{ij}w_i}{w_j}, i,j \in \mathscr{N}$, where $w = (w_1, \cdots, w_n)$, we obtain a simplified form of the linear fractional function on $\mathbf{R}^n_+$, which is a quotient of two linear functions. Function (18) is not convex, it is, however, quasi-convex. More precisely, it is strictly quasi-convex on $\mathbf{R}^n_+$, the positive orthant of $\mathbf{R}^n$. Hence, $\mathscr{E}_A(w) = \max\{\max\{\frac{a_{ij}w_j}{w_i}, \frac{w_i}{a_{ij}w_j}\} | i,j \in \mathscr{N}\}$ is strictly quasi-convex on $\mathbf{R}^n_+$.

It is a well-known fact saying that strictly quasi-convex functions are unimodal, i.e. each local minimum of a strictly quasi-convex function is a global minimum (see e.g. [2]). Summarizing the above stated facts, we obtain that the objective function is unimodal. Taking into account that constraints (14), (16) in (Problem ε), i.e. variant (II), define a convex set, we conclude that the set of all optimal solutions of (Problem ε), variant (II), is convex and each local optimal solution is global. Consequently, by solving (Problem ε), variant (II), e.g., by some interior point method (see [2]), we arrive at the global optimal solution.

Alternatively, variant (II) of (Problem ε) can be solved by a sequence of linear problems as follows (so called the epigraph method, [2]). Instead of minimizing objective (18) subject to constraints (15), (16), we set $t = \frac{a_{ij}w_i}{w_j}$ and solve the system of linear constraints in each iteration.

Constraints (17) in (Problem ε), however, need a special treatment. The set of vectors $w = (w_1, \cdots, w_n)$ fulfilling constraints (17) are not convex, and therefore, the usual interior point methods for solving the optimization problem (18), (15) - (17) could be inefficient, or, fail. That is why we propose for solving (Problem ε), variant (III), the popular penalty methods, see e.g. [2]. Moreover, the constraints in (Problem ε) can be also (partly) linearized by substituting new variables, $y_{ij} = \frac{w_i}{w_j}, i,j \in \mathscr{N}$.

5 Conclusion

In this paper we investigated some important and natural properties of PCMs called the desirable properties, particularly, the non-dominance, consistency, intensity and coherence, which influence the generated priority vectors for final ranking of the given alternatives. The purpose is to calculate the priority vector characterizing the ranking of elements for non-consistent pairwise comparisons matrices. There exist various methods for calculating the vector of weights, e.g. Saaty's Eigenvector Method, the Arithmetic Mean Method, Geometric Mean Method, Lest Square Method and others. We proposed newly reformulated desirable properties – the non-dominance, consistency, intensity, and coherence – of the priority vector, investigated their properties and we also proposed a new method based on an optimization problem how to generate priority vectors with these desirable properties of the given pairwise comparison matrix.

Acknowledgements: This research has been supported by GACR project No. 21-03085S.

References

[1] Bana e Costa, C.A. & Vansnick, J. C. (2008). A critical analysis of the eigenvalue method used to derive priorities in the AHP. *European Journal of Operational Research* 187, 3, 1422–1428.

[2] Boyd, S. & Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paolo, Delhi.

[3] Kulakowski, K., Mazurek, J., Ramik, J. & Soltys, M. (2019). When is the condition of order preservation met? *European Journal of Operational Research* 277, 248–254.

[4] Ramik, J. (2017). Ranking Alternatives by Pairwise Comparisons Matrix and Priority Vector. *Scientific Annals of Economics and Business* 64, 85–95.

[5] Ramik, J. (2020). *Pairwise comparisons method: Theory and Applications in Decision Making.* Switzerland, Cham—Heidelberg—New York—Dordrecht—London: Springer Internat. Publ. 253 p.

[6] Saaty, T.L. (1977). A Scaling Method for Priorities in Hierarchical Structure. *Journal of Mathematical Psychology* 15, 3, 234–281.

Unveiling the Myth: Investigating the Existence of Hot Hands in Gaming

Jan Rejthar[1]

Abstract. This paper addresses persistent issues in the literature surrounding the Hot Hand phenomenon, which has attracted research attention in sports, betting, and investing for decades. The study utilizes data from players of KovaaK's aim trainer as a unique approach to overcome limitations faced by the existing literature. Rigorous Monte Carlo permutation tests are conducted on the data, controlling for family-wise error rate (FWER) and setting a false discovery rate (FDR), to investigate the existence of the Hot Hand. Surprisingly, the results indicate no evidence to support the notion of a Hot Hand in the KovaaK's data. The study concludes that gamers' hands do not possess the mythical Hot Hand ability.

Keywords: Hot Hand, permutation tests, gaming, KovaaK's

JEL Classification: C12, C18, Z29
AMS Classification: 91E45

1 Introduction

The Hot Hand phenomenon is characterized by a heightened probability of success following a sequence of consecutive successful attempts and can be viewed as a manifestation of psychological momentum, which has been viewed as an important contributor to success [6]. It has been a staple part of sports lexicon for a long time and it has received considerable attention in the academic world ever since [5], [10] and [11] have concluded that the Hot Hand does not exist and is just a fallacy stemming from misperception of random sequences. The apparent non-existence has been considered as a possible explanation of various psychological and economic phenomena [7]. While results have been mixed, newer research is arguably more in favor of The Hot Hand being real [3].

Generally, the existing literature suffers from 6 major shortcomings. Firstly, most papers use data from sports, such as basketball, football, or tennis, where players interact with each other all the time, which is a source of complex endogeneity. Secondly, heterogeneity of attempts is often ignored. Thirdly, a lot of papers suffer from short strings of attempts, due to the disciplines analyzed, which is a limiting factor for a lot of methods. Fourthly, time dimension tends to be ignored, which may overlook a "cooling" effect. Fifthly, most commonly used methods lack statistical power and are thus unable to detect the hot hand even if present. And lastly, most papers suffer from a small sample bias [1], [2], [3], [6] and [7]. The resolution of all the aforementioned issues has solely been achieved through controlled experiments.

2 Data

This paper utilizes a unique dataset of KovaaK's players that circumvents the data limitations encountered in previous research without the downsides associated with controlled experiments. KovaaK's is an aim trainer designed to enhance players' aiming proficiency in shooter games and can be used similarly to how drills are used in sports. There is no player interaction in KovaaK's, which was emphasized by [8] as a great advantage of using darts and by [4] as one of the advantages of using bowling. This eliminates many sources of endogeneity.

KovaaK's differs from disciplines such as basketball, football, darts, or bowling in that it allows for longer strings of shots, with most sequences surpassing 100. Although KovaaK's does not record the time intervals between shots, the duration of each scenario is brief, either 60 or 30 seconds. This means that even the shortest sequence of 64 or 82 shots respectively has an average interval of less than 1 second per shot in each round. Additionally, while some individual shots in certain scenarios may be slightly more challenging than others, the variations are insignificant enough that the shots can be considered homogeneous.

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague, Czech Republic, rejj02@vse.cz.

Players were asked to upload their KoovaK's data files anonymously and voluntarily via a Dropbox link. A total of 41,756 runs were gathered, and out of these, 684 runs were chosen for the analysis. The selected runs were then separated into two sub-samples depending on the ability to plan ahead. The first sub-sample consisted of 24 runs where planning ahead was not feasible since only one target was visible at a time. The second sub-sample comprised of 660 runs where planning ahead was possible as three targets were visible simultaneously. Table 1 provides summary statistics of the sub-samples.

Type	N	Mean	Median	Sd	Min	Max
Without planning	24	71.800	72.000	3.784	64.000	80.000
With planning	660	136.109	137.000	10.606	82.000	190.000

Table 1 Summary statistics of both sub-samples

3 Methodology

This paper follows the methodology of [7] and [9] and tests whether there is no positive serial dependence in the outcomes of observed shots. Given a vector of shot outcomes X_i of a run i which takes on value 1 for a hit and 0 for a miss, this paper tests the null hypothesis:

H_0: X_i is i.i.d.,

for all runs with the assumption that the shots came from a Bernoulli processes. Following [5], a player can be viewed as having the hot hand, if in a run i the probability they hit after a streak of k consecutive hits starting at shot j is higher than the marginal probability of them hitting, that is

$$t_P^k(P_i) = P_i\left\{X_{i,j+k} = 1 | \prod_{l=0}^{k-1} X_{i,j+l} = 1\right\} - P_i\{X_{i,j+k} = 1\} > 0 \tag{1}$$

or if the probability they hit after a streak of k consecutive hits starting at shot j is higher than the probability they hit after a streak of k consecutive misses starting at shot m, that is

$$t_D^k(P_i) = P_i\left\{X_{i,j+k} = 1 | \prod_{l=0}^{k-1} X_{i,j+l} = 1\right\} - P_i\left\{X_{i,j+k} = 1 | \prod_{l=0}^{k-1}(1 - X_{i,m+l}) = 1\right\} > 0. \tag{2}$$

Empirical probabilities (proportions) can be utilized to estimate the statistics $t_P^k(P_i)$ and $t_D^k(P_i)$. The difference between the proportion of hits after k consecutive hits and the overall proportion of hits in a run i of length n can be represented by $\hat{P}_{n,k}(X_i) - \hat{p}_{n,i}$. Similarly, the difference between the proportion of hits after k consecutive hits and the proportion of hits after k consecutive misses in a run i of length n can be denoted by $\hat{D}_{n,k}(X_i)$.

Statistical inference requires the knowledge of the distributions of $\hat{P}_{n,k}(X_i) - \hat{p}_{n,i}$ and $\hat{D}_{n,k}(X_i)$ statistics, however, those are unknown, and thus need to be simulated. Monte Carlo permutation tests with 100,000 repetitions for each run were performed as a compromise between granularity and computation intensity. Since multiple runs were analyzed, simultaneous inference needs to be addressed. Two controls were applied: Family-wise error rate (FWER) using the step-down Holm-Šidák procedure and False discovery rate (FDR) using the Benjamini-Hochberg procedure.

4 Results

The results of the permutation tests can be found in Table 2. The table shows the number of individual H_0 that were rejected at 0.05 when tested separately and when simultaneous inference was controlled for. In the case of the sub-sample with possible planning, there are numerous rejected hypotheses when tested separately. For

instance, 56 out of the 660 runs were found to differ from randomness for $k = 2$ at the 0.05. However, none of the hypotheses were rejected once simultaneous inference was considered and corrected for by either FWER or FDR.

In the case of the sub-sample without planning, only one individual H_0 was rejected for $k = 2$ when the $\widehat{D}_{n,k}(X_i)$ estimator was used, which did not survive simultaneous inference either. Putting the results together, the evidence goes starkly against the existence of the Hot Hand in the case of KovaaK's players.

With planning					Without planning						
		$\hat{P}_{n,k}(X_i) - \hat{p}_{n,i}$				$\widehat{D}_{n,k}(X_i)$			$\hat{P}_{n,k}(X_i) - \hat{p}_{n,i}$		
k	N	Individual	Simultaneous		N	Individual	Simultaneous		Individual	Simultaneous	
			FWER	FDR			FWER	FDR		FWER	FDR
2	660	56	0	0	24	1	0	0	0	0	0
3	660	37	0	0	15	0	0	0	0	0	0
4	660	45	0	0	7	0	0	0	0	0	0
5	660	36	0	0	2	0	0	0	0	0	0
6	660	29	0	0	1	0	-	-	-	-	-
7	660	28	0	0							
8	660	30	0	0							
9	660	28	0	0							
10	657	26	0	0							

The table shows the number of hypotheses rejected at 0.05. k is the number of consecutive hits to have the hot hand and N is the number of runs tested.

Table 2 Results of the permutation tests.

5 Results

This paper concludes that KovaaK's players do not experience the Hot Hand. While it is difficult to extrapolate this finding and assert that the Hot Hand is only a cognitive misperception, it at least suggests that psychological momentum might not exist in every context. Additionally, this paper argues that the KovaaK's data used are the most robust non-experimental data used in the Hot Hand literature.

Acknowledgements

This work was supported by The Internal Grant Agency of Prague University of Economics and Business under Grant VŠE IGS F4/52/2023.

References

[1] Avugos, S., Bar-Eli, M., Ritov, I. & Sher, E. (2013). The elusive reality of efficacy–performance cycles in basketball shooting: An analysis of players' performance under invariant conditions. *International Journal of Sport and Exercise Psychology*, *11*(2), 184–202. https://doi.org/10.1080/1612197X.2013.773661

[2] Bar-Eli, M., Avugos, S. & Raab, M. (2006). Twenty years of "hot hand" research: Review and critique. *Psychology of Sport and Exercise*, *7*(6), 525–553. https://doi.org/10.1016/j.psychsport.2006.03.001

[3] Bar-Eli, M., Krumer, A. & Morgulev, E. (2020). Ask not what economics can do for sports-Ask what sports can do for economics. In *Journal of Behavioral and Experimental Economics*, 89, 101597. https://doi.org/10.1016/j.socec.2020.101597

[4] Dorsey-Palmateer, R. & Smith, G. (2004). Bowlers' hot hands. *The American Statistician*, *58*(1), 38–45. https://doi.org/10.1198/0003130042809

[5] Gilovich, T., Vallone, R. & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

[6] Iso-Ahola, S. E. & Dotson, C. O. (2014). Psychological Momentum: Why Success Breeds Success. *Review of General Psychology*, *18*(1), 19–33. https://doi.org/10.1037/a0036406

[7] Miller, J. B. & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, *86*(6), 2019–2047. https://doi.org/10.3982/ECTA14943

[8] Ötting, M., Langrock, R., Deutscher, C. & Leos-Barajas, V. (2020). The hot hand in professional darts. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *183*(2), 565–580. https://doi.org/10.1111/rssa.12527

[9] Ritzwoller, D. M. & Romano, J. P. (2022). Uncertainty in the hot hand fallacy: Detecting streaky alternatives to random bernoulli sequences. *The Review of Economic Studies*, *89*(2), 976–1007. https://doi.org/10.1093/restud/rdab020

[10] Tversky, A. & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance*, *2*(1), 16–21. https://doi.org/10.1080/09332480.1989.11882320

[11] Tversky, A. & Gilovich, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance*, *2*(4), 31–34. https://doi.org/10.1080/09332480.1989.10554951

The Impact of the COVID-19 Pandemic in the Brewing Industry with Regard to Profitability, Cost and Production Efficiency

Jana Sekničková[1], Martina Kuncová[2]

Abstract. A large number of economic subjects in the Czech Republic are firms operating in the manufacturing industry. Most of these firms transform inputs such as labour, capital and raw materials into outputs such as products. Most of these firms then evaluate the efficiency of their operations in their annual report. The efficiency analysis of a firm or its branches is usually carried out using ratio indicators. However, modern approaches allow the use of more sophisticated tools for analysis, such as data envelopment analysis (DEA) models, which allow firms to be evaluated more comprehensively and to include a greater variety of inputs and outputs without having to explicitly specify the relationships between them.. Brewing companies have been evaluating their efficiency in the same way for many years. However, the COVID-19 pandemic has had a significant impact on the foodservice sector as a whole, significantly affecting not only firms' profits but also consumption and, consequently, beer production. A lot of companies have been forced out of business by the pandemic. This paper focuses on the evaluation of firms in the brewing sector with respect to profit, cost and production efficiency and examines the impact of the COVID-19 pandemic in the Czech Republic on the whole brewing sector in order to analyze the main sources of inefficiency. The results of the DEA models used showed that of the 124 breweries studied, only 2 were efficient in all 3 perspectives (profit, cost and production efficiency) and furthermore the hypothesis that profit efficient firms are also production and cost efficient was not confirmed in the sector.

Keywords: DEA model, brewing industry, profitability, production efficiency

JEL Classification: C44, C61, C67
AMS Classification: 90B50, 90C08

1 Introduction

The Czech Republic has a rich history of beer production, with records of brewing dating back to the early first millennium. The earliest mention of beer production in the region can be traced back to around 1088 [8],[2]. The growth of brewing, alongside existing monastic breweries, is closely tied to the establishment of royal towns, particularly in the 12th and 13th centuries. However, in the late 19th and early 20th centuries, there was a significant decline in the number of breweries. According to reports from the Provincial Statistical Office [8], the number of breweries decreased from approximately 1000 in 1841 to half that number in 1925. This trend led to the consolidation of production in larger breweries. The decline continued until the 1960s, when the number of breweries dropped below 100. Since then, the number of industrial breweries, which produce over 10,000 hectolitres per year, has remained relatively stable. In 1999, there were 55 industrial breweries in the Czech Republic, and by 2019, the number had slightly decreased to 52. On the other hand, the number of craft breweries experienced significant growth, increasing from 27 in 1999 to 450 in 2019, as reported by Tripes [20] and Tripes and Dvořák [21]. Presently, the Czech Republic has nearly 600 breweries, positioning it as the seventh country in Europe with the highest number of active breweries, according to Statista.com [17].

The main focus of breweries is beer production. The highest volume of beer production in the Czech Republic was reached in 2019, when beer production exceeded 21.5 million hectoliters. The most popular are lagers, i.e. bottom-fermented 11s and 12s, which account for more than half of the total domestic production [7].

[1] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, jana.seknickova@vse.cz
[2] Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, kuncovam@vse.cz

This article focuses on the efficiency measurement of Czech breweries in 2018 - 2020 using DEA models. The aim is to assess the cost, production and profit efficiency of selected Czech breweries.

2 Efficiency Measurements

The analysis of firm behaviour belongs to the basic microeconomic models and necessarily becomes the subject of optimization. From a microeconomic point of view, a firm's behaviour can be considered optimal if:

- maximizes output at a bounded (specified) cost;
- minimizes costs for a bounded (fixed) level of output; and
- maximizes profit.

Intuitively, in economic terms, we consider firms that achieve the maximum possible revenue at the minimum possible cost to be efficient. This then directly implies profit maximization and all three of the previously mentioned approaches relates to rational firm behavior.

Of course, there are many different approaches to evaluating firms and their efficiency and performance [18]. However, practically all of them are based on the assumption that a firm behaves efficiently if it consumes the minimum amount of inputs (e.g., labour and capital) to ensure the maximum production of outputs. Multi-criteria decision-making methods (inputs are maximized and outputs are minimized), ratio analysis (in particular, maximizing outputs per unit of input) or data envelopment analysis models can be used to evaluate the efficiency of firms. However, a fundamental problem in all these approaches is the correct choice of relevant inputs and outputs. There is certainly no universal approach to determine which characteristics should be used as inputs and which as outputs.

2.1 Inputs

In manufacturing companies, it is possible to build on experience and studies and generally consider as inputs the minimization indicators that need to be provided before production and whose quantity can be influenced by the company's decision-making. Therefore, inputs typically include raw materials for production, the capital and labour required, production machinery and production space [6].

However, not all inputs are of a minimizing type, and typical inputs may also be, for example, pre-contracted orders. Clearly, firms that behave efficiently are likely to have more such orders than inefficient firms. Hence, the number of orders is a desirable input, and we want to maximize rather than minimize it.

2.2 Outputs

Outputs in the manufacturing sector are generally considered to be maximization indicators related to direct production, such as the number of products produced, production volumes, or to subsequent sales, such as the number of products sold, sales revenue or total profit [5]. These indicators are usually not directly influenced by the firm and their value is influenced by changes in inputs or transformation processes, in particular by adjusting technological processes or purchasing new more efficient machines, etc.

But not all outputs are maximizing in nature. The production process can generate emissions, unusable waste and other unnecessary or even harmful components that we consider undesirable outputs. As a rule, we then try to minimize the values of such indicators.

2.3 Efficiency

Every company that produces products or services has to deal with the issue of efficiency and performance. These entities work with limited resources and their main objective should be to optimize the use of resources to achieve the highest possible output. There are several possibilities how to measure the efficiency or performance of a company. Tangen [18] mentioned primarily financial measurements such as Return on sales (ROS), Return on assets (ROA), Return on equity (ROE), or Activity Based Costing (ABC), but as he stated, traditional productivity indicators are also important for monitoring efficiency, either partial or total productivity measures. When considering mathematical-economic approaches, the typical situation involves comparing inputs and outputs. This can be exemplified by cost-benefit analysis, where the costs and benefits associated with a company's production activities are evaluated [15] or data envelopment analysis [6] evaluating different types of company efficiencies. To conduct such analyses, it is essential to accurately identify the specific parameters used. Typically, these parameters are quantitative in nature, such as the number of employees and their salaries (for labour cost

determination), the quantity of products and their prices (to represent sales), financial aspects such as costs, revenues, and profit levels, among others.

Based on a review of existing studies ([3], [5], [6], [9], [13]), three types of efficiencies were selected, which are cost, production and profit efficiencies.

Profit efficiency

Since profit generation is the primary business objective of almost every firm, the profit approach is included as the first aspect in the analysis. The inputs are the indicators that contribute to profit generation, namely employee costs and other operating costs such as purchases of raw materials, equipment, minor repairs, etc. On the output side, we consider the profit of the firm.

Cost efficiency

In cost efficiency, we consider all costs related to production activities on the input side, e.g. labor costs (some studies considering the production approach state that employee costs can account for up to 75% of the total operating costs of a given company or branch), costs of raw materials, machinery, advertising, premises, but also interest paid, etc. The manufacturing firm tries to minimize these costs. The aim is, on the other hand, to maximize the desired financial indicators, which are all revenues, sales of own products, sales of goods sold, etc.

Production efficiency

The production approach is one of the traditional views of evaluating the efficiency of production units, e.g. manufacturing firms operating in an identical market. On the input side, many studies ([9], [14], [16]) have included the number of employees in the management, production and support structure. Other traditional inputs include numbers of production devices, quantities of raw materials consumed, etc.

A less typical input may be the number of customers, which is considered a desirable input. The reason for including this indicator as a desirable input in research is to take into account the fact that a higher number of customers represents a greater awareness of the product and therefore a source of cheap advertising. Advertising costs can then be significantly lower, which of course has the desirable effect of reducing costs and consequently increasing profits. The outputs are the average annual values of the usual production indicators, in particular production volumes, numbers of products sold, etc. A typical output of the production model may also be the number of orders fulfilled.

3 Models and Data

For the analysis, we used a dataset containing economic data from 2018 to 2020 for 124 different firms (102, 98 and 77 in each year of analysis). Data for 215 economic indicators are available, 191 of which could be used to evaluate the efficiency of those firms. Models with a large number of inputs and outputs, however, are not very useful in practice. Although they provide a large amount of information, the interpretation of the results obtained is not very useful. Indeed, on real data, a large number of variables leads to higher values of other variables by which the inputs or outputs need to be adjusted to make the enterprise efficient. In practice, however, in many cases it is not possible to achieve such values or to adjust the data for many variables at once and the results of the analysis are then of little use.

Therefore, and based on the previous efficiency descriptions, we have decided to evaluate firms according to a few inputs and outputs. We evaluated the breweries in each of the 3 years under review in terms of profitability, cost and production efficiency. Since we only had financial indicators (not indicators on total production), we ended up choosing a total of 8 inputs and 3 outputs for the 3 types of models (see Table 1). It is not typical to have only 1 output but here it has several advantages. Using one output per model makes the results more interpretable and easier to communicate. From multiple DEA models, each focusing on a different output, we can conduct comparative analysis among the DMUs based on different dimensions of efficiency (profit, cost, production). The reason for this is also the subsequent possibility to use the same parameters in ratio analysis and compare the results (which will be part of another paper). Last but not least, other researchers have also used only 1 output when analysing breweries [3].

	Inputs	Output
Model A	I1: Personnel costs	O1: Profit/loss
	I2: External resources	
	I3: Equity	
Model B	I4: Labour costs	O2: Total revenue
	I5: Cost of sales + power consumption	
Model C	I6: Number of employees	O3: Revenue from sales of own products and services
	I7: Total liabilities	
	I8: Inventory	

Table 1 List of inputs and outputs for DEA models

Model A represents profit efficiency, model B cost efficiency and model C production efficiency measurements. Table 2 shows the average values for all inputs and outputs considered, calculated across all breweries analyzed in a given year.

To further assess efficiency, we used Data envelopment analysis (DEA) models, specifically input-oriented models with both constant and variable returns to scale.

	Year	**2018**	**2019**	**2020**	
	Number of companies	102	98	77	**model**
CZK thousands	I1: Personnel costs	41 304	46 008	56 876	A
	I2: External resources	261 662	166 187	260 467	A
	I3: Equity	266 517	386 812	442 874	A
	I4: Labour costs	29 394	33 238	40 843	B
	I5: Cost of sales + power consumption	181 194	192 742	229 231	B
x	I6: Number of emplyees	57	55	7	C
CZK thousands	I7: Total liabilities	529 704	553 859	704 197	C
	I8: Inventory	27 010	26 633	36 189	C
	O1: Profit/loss	25 378	63 971	67 747	A
	O2: Total revenue	339 807	362 581	420 184	B
	O3: Revenue from sales of own products and services	308 684	329 650	377 301	C

Table 2 Average values for all inputs and outputs of all breweries

3.1 DEA Model

Traditionally, the input-oriented model is chosen for analysing the behaviour of manufacturing firms, as the manufacturing firm can influence inputs rather than outputs to achieve efficiency.

Thus, let us consider a set of n homogeneous production units $U_1, U_2, \ldots, U_n$ (in general, units $U_i, i = 1, 2, \ldots, n$), which we evaluate based on m inputs and r outputs. Denote by $\mathbf{X} = \{x_{ij}, i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m\}$ the matrix of inputs and similarly by $\mathbf{Y} = \{y_{ik}, i = 1, 2, \ldots, n,\ k = 1, 2, \ldots, r\}$ the matrix of outputs. We can then determine the efficiency of the q-th unit U_q in a given set by solving the basic CCR-I model [4] in the case of constant returns to scale (CRS) or the BCC-I model [1] in the case of variable returns to scale (VRS). For practical computations the original models which correspond to linear fracture programming problems are transformed by the Chranes-Cooper transformation into linear programming problems and then dual problems are formulated for them. The optimization problem can then be written in the general form (1):

min

$$\theta_q - \varepsilon\left(\sum_{k=1}^{r} s_k^+ + \sum_{j=1}^{m} s_j^-\right),$$

subject to

$$\begin{aligned}
\sum_{i=1}^{n} x_{ij}\lambda_i + s_j^- &= \theta_q x_{qj}, \qquad j = 1, \dots, m, \\
\sum_{i=1}^{n} y_{ik}\lambda_i - s_k^+ &= y_{qk}, \qquad k = 1, \dots, r, \\
\sum_{i=1}^{n} \lambda_i &= 1, \\
\lambda_i, &\geq 0, \qquad i = 1, 2, \dots, n, \\
s_j^- &\geq 0, \qquad j = 1, 2, \dots, m, \\
s_k^+ &\geq 0, \qquad k = 1, 2, \dots, r,
\end{aligned} \tag{1}$$

where

θ_q	denotes the efficiency measure of the evaluated unit U_q,
λ_i	denotes the weight of unit U_i in the dataset,
x_{ij}	denotes the value of the *j*-th input of unit U_i,
y_{ik}	denotes the value of the *k*-th output of unit U_i,
s_j^-	denotes the additive variable of the *j*-th input,
s_k^+	denotes the additive variable of the *k*-th output and
ε	denotes the infinitesimal constant.

The formulation of these models (which we call radial), their transformations and other derived models, including the interpretation of the obtained values, are described in considerable detail in [10]. With respect to returns to scale the two dual models differ from each other only by the convexity condition, i.e. the condition that the sum of the weights of the units in the set must (or in the case of the CCR-I model need not) equal to one.

For models that are referred to as slack-based measure (SBM) models, there is no need to distinguish between model orientations, as they measure efficiency directly using the values of the additive variables. The first SBM model, or additive model, was formulated by Charnes et al. [5]. The formulation of this model can also be found in [6].

The disadvantage of the SBM model is that the efficiency measure is not independent of the change in the scale used for inputs and outputs. Therefore, Tone [19] proposed a model that is also deviation-based, i.e., it measures efficiency based on the additive variables $s_k^+,\ k = 1, 2, \dots, r$ and $s_j^-,\ j = 1, 2, \dots, m$. However, it also satisfies the conditions that the efficiency measure is independent of the units representing the size of inputs and outputs and is a monotonically decreasing function of all additive variables associated with inputs and outputs. According to its author, it is sometimes referred to as the SBMT model. Such a model also appears to be appropriate for evaluating the efficiency of manufacturing firms. The model is similar to the model (1), only the objective function and the first set of constraints have the following form (2):

min

$$\rho_q = \frac{1 - \frac{1}{m}\sum_{j=1}^{m}\left(\frac{s_j^-}{x_{qj}}\right)}{1 + \frac{1}{r}\sum_{k=1}^{r}\left(\frac{s_k^+}{y_{qk}}\right)}, \tag{2}$$

subject to

$$\sum_{i=1}^{n} x_{ij}\lambda_i + s_j^- = x_{qj}, \qquad j = 1, \dots, m.$$

3.2 DEA and Breweries Efficiency

Measuring brewery efficiency using DEA models appears in several articles. Bernetti et al. [3] conducted an efficiency analysis of 163 Italian microbreweries. They employed input-oriented DEA models with three inputs (number of employees, debt/equity ratio, total debt) and one output (revenues from sales and services), considering both CRS and VRS models. Sellers Rubio [14] focused on estimating advertising efficiency in the Spanish beer industry. They utilized input-oriented DEA models to analyze a sample of six beer firms that operated continuously from 2007 to 2014. To calculate advertising efficiency, they considered six inputs (four different advertising expenditures, number of employees, capital) and two outputs (total sales revenues, total beer sales). Ezan [9] employed a two-staged model to measure the DEA efficiency of 500 beer production companies. The first stage involved profitability analysis, where the output variables (revenue and profits from operations) served as inputs in the second stage, known as marketability analysis. In the second stage, three input variables (total assets excluding financial investments and investment properties, stockholders' equity, and total number of employees) were considered, while four indicators (earnings per share, average stock price, return on invested capital, net income) were used as outputs. In addition to DEA models, Solomon [16] utilized econometric models, specifically stochastic frontier production function models, to analyze the efficiency of breweries.

Czech breweries were also subject to efficiency analyses. Kasem et al. [11]conducted research on Czech breweries, focusing on efficiency assessment. They utilized indicators from the Global Reporting Initiative and Key Performance Indicators, along with environmental, financial, and sustainability report data sets. The study involved 14 breweries and employed DEA models. The DEA models used four inputs: number of employees, average employees' salary and bonus, amount of waste, and percentage of women in supervisory positions within the company. Two outputs were considered: cash flow and economic value added (EVA). In another study by Kasem and Trenz [12], a three-phase system was proposed for assessing the sustainability of 89 Czech brewery companies. The first phase involved data collection and the calculation of Sustainability Value Added (SVA) based on concepts of environmental, social, and governance value added. The second phase included DEA analysis and the sustainability assessment model. To automate the processes in the first two phases, the authors utilized the web portal Web Information System for Corporate Performance Evaluation and Sustainability Reporting (WEBRIS) for data extraction. Lastly, the analysis was conducted based on questionnaires and data from the 89 Czech breweries.

4 Results

In our analysis, we calculated measures of three efficiencies - cost, production and profit - for 124 breweries using input-oriented DEA models. Based on previous studies, the variable returns to scale (BCC-I) models proved to be the most appropriate. For profit efficiency, we chose as inputs personnel costs, external resources and equity; as output, profit or loss was selected. For cost efficiency, we chose labor costs, cost of sales and power consumption as inputs; total revenue was chosen as output. Finally, for production efficiency, we chose as inputs the number of employees (increased by 1 to be able to cover small companies without employees with 1 owner), total liabilities, and inventories; as outputs, revenue from sales of own products and services chosen.

4.1 Profit Efficiency

With profit maximization as the main economic objective, we expected high levels of profit efficiency in breweries. However, average profit efficiency ratios were surprisingly low and also the number of efficient firms

was low. In 2018 - 2020, only 13.7%, 10.2% and 18.2% of firms were assessed as efficient by profit efficiency, respectively. The average efficiency rate was around 24.9% in all years. Over the full time period, only 4 firms out of 124 (3%) were profit efficient. This is surprising as the profit maximization is the main economic objective of any manufacturing firm.

It should be noted that only 56 of the 124 firms surveyed had been doing business on the Czech market for the full three years. We attribute the closure or the establishment of new breweries to the coronavirus crisis.

4.2 Cost Efficiency

Cost minimization is also one of the basic economic approaches and therefore we assumed that many breweries would be cost-effective. Also in this case only 32.4%, 28.4%, and 28.6% of firms were cost-efficient in 2018 - 2020, respectively. 15 of the 124 firms surveyed (12%) were even cost-efficient in all three years. The average cost efficiency was about 50.9%.

4.3 Production Efficiency

During 2018 - 2020, 57.9%, 57.1% and 80.5% of beer producing companies in the Czech Republic were evaluated as production efficient. In all three years, however, only 21 out of 124 enterprises (17%) were production efficient. These firms were therefore efficient in their beer selling behaviour throughout the period under review. Nevertheless, there were also firms in the national network whose relative degree of efficiency fluctuated significantly over the three years of the study. For different firms, we observed a change in the degree of efficiency in all directions. Thus, there were firms in the sample that had very low efficiency rates at the beginning of the study period and were rated efficient in 2020, and similarly, there were firms that were efficient only in 2018, after which their efficiency rates dropped significantly. There were also firms observed that were efficient in the first and third years but showed low efficiency rates in 2019. The average production efficiency rate in all years under study was around 58.8%.

These changes could be partly explained by the coronavirus crisis, which affected all regions in the time period under review and which could have a different impact depending on the location in which the company operates. However, data on the scope of the operations were not available (only data on the firm's headquarters are available), and thus a more detailed analysis of the relationship between the impact of the coronavirus crisis and the firm's location is not possible from this perspective. However, the data shows a significant change in recording or reporting the number of employees during the pandemic restrictions in 2020, with most firms reporting 0 employees.

4.4 Total Efficiency

Let us denote as total efficient those firms that behave efficiently in terms of profit maximization, cost minimization and production maximization throughout the observation period. Thus, total efficient firms are those that are simultaneously profit efficient, cost efficient and production efficient. Of the 124 firms, 7, 8 and 4 breweries were total efficient in 2018 - 2020, i.e. less than 6.5%. The remaining firms were inefficient in some sense.

Over the full time period, from 124 firms evaluated, 6 were overall efficient, but only two of them remained in the market for all three years. These were Plzeňský Prazdroj a.s (the biggest brewery in the Czech Republic) and Mandala CZ, s.r.o. (a small canteen in the Pardubice region in the town of Hlinsko, which has among its business activities the production of beer from malt). We therefore consider these two breweries to be overall efficient.

5 Conclusions

In this paper we have tried to look at modelling efficiency based on purely microeconomic assumptions for the rational behaviour of a manufacturing firm. We measured brewery efficiency using cost, production and profit efficiency.

Let us formulate the hypothesis that firms that behave profit efficiently are also cost and production-efficient. However, this hypothesis has not been confirmed in the brewing industry. None of the correlation coefficients measuring the dependence between the outcomes of the various efficiencies exceeded 34%, and the strongest dependence was between profit and production efficiency. In this case, although the statistical test rejects the

independence hypothesis, the degree of dependence is much weaker than would be expected from economic theory.

Based on the available data, we conducted an analysis and concluded that out of 124 brewing companies doing business in the Czech Republic, only 2 behave efficiently in all senses throughout the period under study. They are Plzeňský Prazdroj, a.s. and Mandala CZ, s.r.o.

Acknowledgements

The research was supported by an institutional fund IP400040 for long-term conceptual development of science and research at the Faculty of Informatics and Statistics, Prague University of Economics and Business and also by the Internal Grant Agency of the Prague University of Economics and Business IGA F4/42/2021.

References

[1] 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.
[2] Basařová, V. (1999). *Základní charakteristika a principy výroby piva.* České pivo, Praha: NUGA.
[3] Bernetti, I., Alampi Sottini, V., Cipollaro, M. & Menghini, S. (2020). A survey on the performance of the Italian brewing companies. *Italian Review of Agricultural Economics*, 75(1), 37–50.
[4] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2(6), 429–444.
[5] 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, 91–107.
[6] Cook, W. D. & Seiford, L. M. (2009). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research*, 192, 1–17.
[7] Czech Statistical Office (2019). *100 years of statistics*, [Online]. Available at: https://www.czso.cz/csu/stoletistatistiky/prvni-republika-pivni-republika [cited 2023-03-24].
[8] Danišová, I. (2020). Zlatý mok očima statistiky. *Statistika & My*, 10/2020, 20–21.
[9] Ezan, A.I. (2015). *Efficiency Measurements in the Turkish Brewing Industry by Using Data Envelopment Analysis.* [Online]. Available at: https://kanazawa-u.repo.nii.ac.jp/ [cited 2023-03-26].
[10] Jablonský, J. & Dlouhý, M. (2004). *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing, s.r.o.
[11] Kasem, E., Trenz, O., Hřebíček, J. & Faldík, O. (2015). Key Sustainability Performance Indicator Analysis for Czech Breweries. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 63(6), 1937–1944.
[12] Kasem, O. & Trenz, O. (2020). Automated Sustainability Assessment System for Small and Medium Enterprises Reporting. *Sustainability*, 2020, 12, 5687, 1–23.
[13] Sahoo, B. K., Mehdiloozad, M., & Tone, K. (2014). Cost, revenue and profit efficiency measurement in DEA: A directional distance function approach. *European journal of operational research*, 237(3), 921–931.
[14] Sellers Rubio, R. (2018). Advertising efficiency in the Spanish beer industry: spending too much?, *International Journal of Wine Business Research*, 30(4), 410–427.
[15] Smith, P. C., & Street, A. (2005). Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 168(2), 401–417.
[16] Solomon, N. (2018). *Technical Efficiency Analysis of the Ethiopian Brewery Industries.* [Online]. Available at: https://nadre.ethernet.edu.et/record/8752#.ZAJqMB-ZNPY [cited 2023-03-23].
[17] Statista.com (2021). *Number of active beer breweries in Europe in 2021, by country,* [Online]. Available at: https://www.statista.com/statistics/444614/european-beer-breweries-by-country/ [cited 2023-03-24].
[18] Tangen, S. (2003). An overview of frequently used performance measures. *Work study*, 52(7), 347–354.
[19] Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research,* 130(3), 498–509.
[20] Tripes, S. (2019). Leaders of Czech Craft Breweries and Their Ability to Attract the Customer. *15th European Conference on Management, Leadership and Governance*, Porto, Portugal, 383–390.
[21] Tripes, S. & Dvořák, J. (2017). Strategic forces in the Czech brewing industry from 1990-2015. *Acta Oeconomica Pragensia,* 25(3), 3–38.

Nowcasting Unemployment Using Mixed Data Sampling and Google Trends Data

Tereza Singerová[1], Lukáš Frýd[2]

Abstract. Forecasting macroeconomic variables commonly encounters the problem that data describing the current state are available with a significant, sometimes annual, lag. One possible solution is to use data from Google Trends for nowcasting the current values of macroeconomic variables. However, Google Trends data is available at a different, usually higher, frequency than macroeconomic data. To avoid losing potentially useful information in the form of aggregating these data, MIDAS regression can be used to link data with different frequencies. Using the Czech Republic's unemployment as an example, we show the higher predictive power of the combination of Google Trends and MIDAS regression compared to the traditionally used ARIMA and ARIMAX models.

Keywords: Google Trends, MIDAS, Nowcasting, Unemployment

JEL Classification: C53, E24
AMS Classification: 62P20

1 Introduction

Econometric time series forecasting methods heavily rely on autoregressive models. In the field of macroeconomics, vector autoregression is the most commonly utilized approach. However, one challenge we encounter is the frequency of data collection and the delays involved in data processing and publication. Depending on the specific variable, there can be significant lags ranging from weeks to even up to a year for obtaining the latest data. Consequently, the resulting predictions are subject to considerable delays.

In this study, our focus is on predicting unemployment in the Czech Republic using quarterly data. With this type of data collection, it is possible to make forecasts with a quarterly lag. Nonetheless, to bridge the gap between the currently available published data and the upcoming data collection, one potential solution is to leverage Google Trends data. By analyzing the trend in keyword searches, we can approximate unemployment trends. As a result, we can make predictions with a higher frequency compared to waiting for updated official data. This forecasting approach is known as nowcasting. Specifically, we employ nowcasting techniques to predict the unemployment rate in the Czech Republic using Google Trends statistics. Changes in the search volume of terms related to unemployment can serve as signals for changes in the unemployment rate. Given the disparity in data collection frequency between official data and Google Trends data, we employ the Mixed-data sampling (MIDAS) model. The MIDAS model enables the combination of data collected at different frequencies, ensuring that no information is lost from higher-frequency data sources. We compare the predictive performance of a model that incorporates both macroeconomic variables and Google Trends data with a traditional ARIMA and ARIMAX model, which solely relies on lagged dependent variables and macroeconomic variables. The results indicate that combining Google Trends data with the MIDAS model yields superior predictive capabilities compared to the ARIMA and ARIMAX models.

2 Data

We combine two data sources to compare whether the use of search data improves models that used only macroeconomic data or autoregression.

2.1 Google Trends Data

Google Trends data offers valuable insights into labor and household activity, surpassing surveys in terms of speed, cost-effectiveness, and accuracy in capturing individual or consumer behavior. Several studies have explored the potential of Google Trends data for predicting unemployment rates and improving forecast quality.

[1] University of Economics in Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, sint02@vse.cz
[2] University of Economics in Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, lukas.fryd@vse.cz

Askitas and Zimmermann [1] demonstrated the usefulness of keyword searches in forecasting monthly unemployment rates in Germany using Google Trends data. Vicente et al. [6] combined traditional indicators with search data, specifically focusing on the keyword "job offers," to analyze the Spanish economy using an ARIMA model. Their findings highlighted the enhanced forecasting quality resulting from incorporating search data. Similarly, Smith [5] and Maas [3] utilized the MIDAS model to predict unemployment in the UK and found that Google Trends data significantly improved predictive capabilities. However, Maas [3] also noted a decline in the predictive power of Google trends data over time, emphasizing its effectiveness primarily in the short term. In the context of the Czech Republic, Pavlíček and Krištoufek [4] were the only researchers who explored the relationship between unemployment and Google Trends data. They aggregated high-frequency variables and demonstrated the potential benefits of incorporating internet data into predictive models.

Google Trends provides a volume index representing the relative frequency of keyword searches compared to a peak value of 100. Although the actual search numbers are not displayed, users can observe the temporal changes in the volume index and filter data by region for global or local insights. The availability of regional data depends on the popularity of the specific keyword.

In the case of the Czech Republic, the search behavior related to unemployment exhibits considerable randomness, with numerous irrelevant terms included. Moreover, users often include specific terms such as years or geographical locations in their searches. This presents a challenge when examining individual sets of key expressions, as the removal of unwanted and irrelevant terms often leaves a limited number of remaining expressions. Hence, we manually selected key terms, categorized them into three groups, and created corresponding indexes. A practical approach to classifying search terms related to unemployment is to divide them into three groups based on the searcher's situation. The first group comprises individuals who are at risk of job loss and seek information on termination and employment termination. The second group consists of recently unemployed individuals who search for state support or benefits. The third group includes job seekers actively searching for new employment opportunities and using search terms associated with job hunting.

2.2 Macroeconomic Data

We uses the ILO's [1] quarterly change in unemployment rate as the dependent variable. We have constructed a new dataset that tracks the quarterly changes of macroeconomic variables from 2004 to Q2 2021 in the Czech Republic. The macroeconomic variables we consider are GDP at current prices and Consumer Price Index. Because of the presence of a unit root for all three macroeconomic variables, we work with their logarithmic differences.

3 Methodology

To assess the predictive accuracy of the MIDAS model, we compare it with three alternative models. The first model is an ARIMA model that solely relies on the dependent variable, which in this case is the unemployment rate, as an autoregressive term. The second model is an ARIMAX model, where we augment the ARIMA model with two additional macroeconomic variables as potential regressors. Lastly, we consider the MIDAS model, which incorporates both the macroeconomic variables and Google Trends data. By conducting this comparative analysis, we aim to determine which model outperforms the others in terms of predictive performance.

3.1 ARIMA

Based on the ACF and PACF plots, We chose the ARIMA (5,1,0) model, which has 5 autoregressive terms and one differencing term to achieve stationarity. The model equation is given by 1, where Δy_t denotes the first difference of the unemployment rate.

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + ... + \phi_5 \Delta y_{t-5} + \epsilon_t,\ \epsilon_t \sim iid(0, \sigma^2) \tag{1}$$

3.2 ARIMAX

The ARIMAX model extends the ARIMA model by allowing additional regressors. We include two macroeconomic variables the GDP growth rate ($x^{(1)}$) and the inflation rate ($x^{(2)}$), with possible lags. The model follows the same specification as before: (5,1,0) for the dependent variable. Equation 2 shows the model.

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + ... + \phi_5 \Delta y_{t-5} + \beta_1 x_t^{(1)} + \beta_2 x_{t-1}^{(1)} + \beta_3 x_t^{(2)} + \beta_4 x_{t-1}^{(2)} + \epsilon_t,\ \epsilon_t \sim iid(0, \sigma^2) \tag{2}$$

[1] https://www.ilo.org/global/lang–en/index.htm

3.3 MIDAS

Google Trends data represent high-frequency indicators that can capture the current state of the economy. However, these data need to be linked to the unemployment variable, which is measured at a lower frequency. The mixed-data sampling (MIDAS) model allows us to combine low- and high-frequency data in a single regression. These methods combine a low frequency variable with a high frequency variable. MIDAS model uses different types of functions to weigh the high frequency variable. In the empirical literature, the most common uses are functions of polynomials. The equation representing the MIDAS model can be expressed as follows:

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i y_t + \boldsymbol{\gamma'} \sum_{k=1}^{m} \phi(k;\boldsymbol{\theta}) L_{VF}^k \boldsymbol{x_t} + \epsilon_t,\ \epsilon_t \sim iid(0,\sigma^2) \tag{3}$$

where the L is lag operator and $\phi(k;\boldsymbol{\theta})$ represents general weighting function. The L operator shifts data back by one period, such that $Ly_t = y_{t-1}$. Applying it twice would shift the data back two periods ($L(Ly_t) = L^2 y_t = y_{t-2}$). We denote the delay operator for high frequency variables as L_{VF}. The function $\phi(k;\boldsymbol{\theta})$ determines how much weight each lag k receives. It can have different forms, such as the normalised Beta function or the normalised exponential Almon lag polynomial, see Ghysels et al. [2]. We utilize the normalised exponential Almon lag polynomial define as:

$$\phi(k;\theta_1,\theta_2) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=1}^{m} exp(\theta_1 j + \theta_2 j^2)} \tag{4}$$

where θ_1 and θ_2 are parameters, m represents the number of lags and j is the index of the lag. The estimation of the equation 4 is based on the nonlinear least squares method.

4 Results

The outputs of the MIDAS model estimation from the 4 equation are shown in the Table 4. The optimal delays and parameter values are found based on minimizing the AIC criteria. The optimal values are: $\theta_1 = 1$ and $\theta_2 = 1,75$ for the first index, $\theta_1 = 1$ and $\theta_2 = 1,25$ for the second index, and $\theta_1 = 1,25$ and $\theta_2 = 1,75$ for the third index. We include five lags of the dependent variable, as in the ARIMA model. The GDP growth rate and the inflation rate are significant at the current and lagged values. The resulting residuals and do not exhibit serial correlation and for this reason we consider the presented model to be dynamically complete.

In order to compare the performance of the models, we assess their error statistics (refer to Table 2) and prediction quality. We re-estimate the models using data from 2004 to 2019, and then generate predictions for the next quarter of 2020. We calculate the mean squared error (MSE) by measuring the deviations between the predicted values and the actual values in an out-of-sample analysis. Based on our evaluation, the MIDAS model, which incorporates both economic data and Google Trends data, achieves the best prediction performance. The results can be found in Table 3.

The MIDAS model demonstrates superior forecasting performance compared to the ARIMA models in predicting the tangentiality factor. It achieves an average error of only 0.1 percent units, whereas the ARIMA models exhibit an average error of 0.5 percent units. This outcome aligns with the error statistics presented in Table 2. One possible explanation for the MIDAS model's superior performance is its inclusion of Google Trends data. These data may capture the impact of the pandemic more swiftly and accurately compared to the data used to train the ARIMA models, which rely on pre-pandemic data up to 2019. The incorporation of more up-to-date and pandemic-specific information could contribute to the MIDAS model's enhanced predictive abilities. To delve further into this matter, a subsequent analysis could explore how different training time intervals affect the predictive capabilities of the models. This investigation would shed light on the sensitivity of the models to the temporal scope of the training data and provide valuable insights into their performance under varying conditions.

In our forecasting process, we employ the MIDAS model to predict future periods with missing data. We include Google Trends data up until March 2022 and incorporate current GDP and inflation rate as regressors in the regression analysis. However, it is important to note that the GDP and inflation rate data are outdated, with the most recent data available from the Czech Statistical Office being for the second quarter of 2021. To overcome this limitation and forecast the upcoming three quarters (Q3 and Q4 of 2021, and Q1 of 2022), we utilize the ARIMA model to generate artificial values for each macroeconomic variable. For the current GDP, we apply an ARIMA(4,1,0) process, while for the consumption price indices, we employ an ARIMA(5,1,0) process. These

	Dependent variable: Unemployment rate		
	Estimate	Standard error	
Constant	0.566	0.117	***
AR(1)	0.991	0.079	***
AR(2)	0.033	0.115	
AR(3)	-0.415	0.096	***
AR(4)	0.821	0.191	***
AR(5)	-0.539	0.085	***
$\Delta\text{HDP}_{(t)}$	0.783	0.565	
$\Delta\text{HDP}_{(t-1)}$	-4.146	0.651	***
$\text{Inflation}_{(t)}$	-1.894	1.864	
$\text{Inflation}_{(t-1)}$	-2.869	1.362	*
$\text{Index}_{1(t)}$	0.023	0.012	.
$\text{Index}_{1(t-1)}$	0.140	0.100	
$\text{Index}_{2(t)}$	-0.030	0.009	**
$\text{Index}_{2(t-1)}$	0.229	0.121	.
$\text{Index}_{3(t)}$	0.017	0.006	**
$\text{Index}_{3(t-1)}$	0.229	0.492	
Note:	$^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001		

Table 1 Estimated MIDAS model

Model	RMSE	MAE	MAPE	MASE	AIC	BIC
ARIMA	0.307	0.219	4.14	0.676	44.8	57.5
ARIMAX	0.301	0.223	4.36	0.678	49.8	70.9
MIDAS	0.263	0.205	0.040	0.606	43.5	78.6

Table 2 Error statistics across models

Model	Mean squared error
ARIMA	0.518
ARIMAX	0.518
MIDAS	0.207

Table 3 Comparison of forecast quality for 2020 quarter across models

ARIMA models allow us to generate estimations for the respective variables. By incorporating these artificial values into the MIDAS model, we can subsequently estimate the unemployment rate for the future periods. It is important to note that while this approach allows for forecasting, the accuracy of the predictions may be influenced by the reliance on artificial values generated through the ARIMA models.

Období	ARIMA	ARIMAX	MIDAS
2021 Q3	2.34	2.38	2.31
2021 Q4	2.29	2.30	2.55
2022 Q1	1.93	1.96	2.32

Table 4 Predicted unemployment rate for 2021Q3-2022Q1

All three models have a similar first predictive value (unemployment rate of around 2.3%). However, the Table 4

shows that the ARIMA and ARIMAX models have similar predictions and only capture the downward trend of the unemployment rate. On the other hand, the MIDAS model shows a spike for the second predictive period, which may be caused by a change in the search volume of an index or a combination of them. The MIDAS model also has narrower confidence intervals, which suggests it gives more precise estimates of the predicted values. Indeed, one limitation of this study is that we utilized Google Trends data at a monthly frequency, even though it is possible to access the data at a higher frequency such as weekly or daily. However, the choice of data frequency is constrained by the length of the survey period provided by Google Trends. Daily data is only available for a maximum 90-day period, while weekly data covers a 5-year period. To obtain higher-frequency data over a longer time span, it would be beneficial to explore methods of connecting this data to a more extensive time series. This could be a valuable avenue for future research to enhance the MIDAS model and improve the accuracy of predictions. By finding ways to incorporate longer and more granular time series data, we can potentially capture more fine-grained dynamics and increase the model's forecasting capabilities.

5 Conclusion

The objective of this paper was to estimate the current unemployment rate in the Czech Republic by utilizing data with different collection frequencies. We incorporated macroeconomic data and Google Trends data for nowcasting and developed three Google indexes that track the relative search volumes of specific key phrases associated with unemployment over time in the Czech Republic. The first index focused on phrases related to job insecurity, while the second index captured searches related to the unemployed seeking state financial assistance. The third index reflected searches for phrases related to job hunting and finding new employment opportunities.

To forecast the unemployment rate, we employed three models and compared their performance. The first model was an ARIMA model that solely utilized past values of the unemployment variable. The second model was an ARIMAX model, which included two additional macroeconomic variables: GDP growth rate and inflation rate. The third model was a MIDAS model that combined Google Trends data with the macroeconomic variables to capture high-frequency information. The ARIMA and ARIMAX models exhibited similar predictions and characteristics, with only slight differences. However, the MIDAS model, which incorporated both data sources, emerged as the most accurate predictor. This demonstrates that Google Trends data can enhance traditional models like ARIMA, resulting in improved forecasting capabilities.

Acknowledgements

The work was supported by the Internal Grant Agency of Prague University of Economics and Business under Grant F4/24/2023.

References

[1] Askitas, N. & Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting. *German Council for Social and Economic Data (RatSWD)*, Research Notes No. 41.

[2] Ghysels, E., Sinko, A. & Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26, 53-90.

[3] Maas, B. (2020). Short-term forecasting of the US unemployment rate. *Journal of Forecasting*, 39, 394-411.

[4] Pavlíček, J. & Krištoufek, L. (2015). Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries. *PLOS ONE*, 10, 1-11.

[5] Smith, P. (2016). Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data. *Journal of Forecasting*, 35, 263–284.

[6] Vicente, M., Lopez-Menendez, A. & Perez Suarez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, 132-139.

Average Reward Optimality in Semi-Markov Decision Processes with Costly Interventions

Karel Sladký [1]

Abstract. In this note we consider semi-Markov reward decision processes evolving on finite state spaces. We focus attention on average reward models, i.e. we establish explicit formulas for the growth rate of the total expected reward. In contrast to the standard models we assume that the decision maker can also change the running process by some (costly) intervention. Recall that the result for optimality criteria for the classical Markov decision chains in discrete- and continuous-time setting turn out to be a very specific case of the considered model. The aim is to formulate optimality conditions for semi-Markov models with intervention and present algorithmic procedures for finding optimal solutions.

Keywords: controlled semi-Markov reward processes, long run optimality, intervention of the decision maker.

JEL classification: C44, C61
AMS classification: 90C40, 60J10, 93E20

1 Formulation and Notation

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 $X\{X_n, n = 0, 1, \ldots\}$. The development of the process $Y(t)$ over time governed the decision maker is the following: At time $t = 0$ if $Y(0) = i$ the decision maker selects decision from a finite $\mathcal{A}_i = 1, 2, \ldots, S$ or from an infinite (compact) set $\mathcal{A}_i \equiv [0, K_i] \subset \mathbb{R}$ 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 $\eta_{ij}(a)$. Let $F_{ij}(a, \tau)$ be a non-lattice distribution function (i.e. the discrete probability distribution concentrated on a set of points of the form $a+nh$ where $h > 0$ and $n = 0, +1, -1, +2, -2, ...$) representing the probability $\mathrm{P}(\eta_{ij} \leq \tau)$. We assume that for $\ell = 1, 2, \quad 0 < d_{ij}^{(\ell)} = \int_0^\infty \tau^\ell \, \mathrm{d}F_{ij}(a, \tau) < \infty$ hence also $0 < d_i^{(\ell)} = \sum_{j=1}^N p_{ij}(a) d_{ij}^{(\ell)}(a) < \infty$. Finally, one-stage transition reward $r_{ij} > 0$ will be accrued to transition from state i to state j, and reward rate $r_i(a)$ per unit of time spent in state i is earned. We assume that each $p_{ij}(a)$ and $r_i(a)$ is a continuous function of $a \in \mathcal{A}_i$.

Moreover, since the decision maker has a complete knowledge on the development of the process over time we assume that the decision maker has an option for additional improvement of the system dynamics by paying certain amount, say $c_i(s)$, to guarantee that the system will jump from state i to state s.

If no intervention is applied the development of the considered Markov process over time is the following.
A (Markovian) policy controlling the semi-Markov process Y, say $\pi = (f^0, f^1, \ldots)$, 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 embedded Markov chain X is in state i. Let π^k be a sequence of decision vectors starting at the k-th transition, hence $\pi = (f^0, f^1, \ldots f^{k-1}, \pi^k)$. Policy which selects at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary; $P(f)$ is a transition probability matrix with elements $p_{ij}(f_i)$. Stationary policy $\tilde{\pi}$ is randomized if there exist decision vectors $f^{(1)}, f^{(2)}, \ldots, f^{(m)} \in \mathcal{F}$ and on following policy $\tilde{\pi}$ we select in state i action $f_i^{(j)}$ with a given probability $\kappa_i^{(j)}$ (of course, $\kappa_i^{(j)} \geq 0$ with $\sum_{j=1}^N \kappa_i^{(j)} = 1$ for all $i \in \mathcal{I}$). For details see e.g. [1, 5, 6].

For the (random) reward earned up to time t, say $\xi(t)$ we have $\xi(t) := \left[\int_0^t r_{Y(s)} \mathrm{d}s \; + \sum_{k=0}^{N(t)} r_{Y(\tau_k^-), Y(\tau_k^+)}\right]$, 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, and $v_i(\pi, t) := \mathsf{E}_i^\pi \xi(t)$ denote the expected total reward of the semi-Markov process $Y(t)$ up to time t given its initial state at time $t = 0$ if policy $\pi = f^n$ is followed. Hence $g_i(\pi) = \lim_{t\to\infty} \frac{1}{t} v_i(\pi, t)$ seems to be the natural definition of average reward generated by the considered semi-Markov process.

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, sladky@utia.cas.cz

On the other hand we can very easily calculated reward generated by the embedded Markov chain. Let ξ_n (resp. τ_n) be the cumulative reward obtained (resp. total time spent) in the n first transitions of the considered embedded Markov chain X. Since the process starts in state X_0, $\xi_n = \sum_{k=0}^{n-1}[r_{X_k} \cdot \eta_{X_k,X_{k+1}} + r_{X_k,X_{k+1}}]$ (resp. $\tau_n = \sum_{k=0}^{n-1} \eta_{X_k,X_{k+1}}$). Similarly let $\xi_{(m,n)}$ (resp. $\tau_{(m,n)}$) be reserved for the cumulative (random) reward, obtained (resp. total time spent) from the mth up to the nth transition. Obviously (we tacitly assume that $\xi_{(1,n)}$ starts in state X_1), $\xi_n = r_{X_0} \cdot \eta_{X_0,X_1} + r_{X_0,X_1} + \xi_{(1,n)}, \quad \tau_n = \eta_{X_0,X_1} + \tau_{(1,n)}$.

If the process starts in state i and policy π is followed let $x_i^\pi(n)$ (resp. $\tau_i^\pi(n)$) be the total expected reward (resp. total time spent) in the n next transitions. The growth rate of $x_i^\pi(n)$ (resp. $\tau_i^\pi(n)$) is linear in time, in particular $x_i^\pi(n) = g^\xi \cdot n + o(n)$ (resp. $\tau_i^\pi(n) = g^\tau \cdot n + o(n)$). Let $G_i(\pi, n) := \frac{x_i^\pi(n)}{\tau_i^\pi(n)}$ is the average reward in the n next transitions, so we can conclude that $G_i(\pi, n) \to G_i(\pi) = g^\xi/g^\tau$.

Comparing the average reward generated by $g_i(\pi)$ and $G_i(\pi)$ we can see that $g_i(\pi)$ represents what is usually meant by the expected average reward, that is the expected reward generated by time t. However, $G_i(\pi)$ also represents at least of some sense the average expectation. While $g_i(\pi)$ is clearly more appealing criterion, it turns that it is easier to work with $G_i(\pi)$. Fortunately, it turns out that under certain regenerative condition both criteria are equal. Roughly speaking, a sufficient condition is that for any stationary policy the resulting semi-Markov process is a regenerative process. For more details see e.g.[5] or [6].

The aim of this note is to formulate optimality conditions for semi-Markov models with additional intervention of the decision maker and present algorithmic procedures for finding optimal solutions. To this end using standard policy iteration procedures we find stationary policy yielding maximum average reward of the considered controlled semi-Markov process.

In what follows we show that under some specific conditions (e.g. if the considered Markov chain contains a single class of recurrent states) for $n \to \infty$ the asymptotic value of average reward is the same for the two definitions mentioned above.

2 Analysis of Average Reward Optimality in Semi-Markov Processes

To begin with we focus attention on average reward optimality in semi-Markov processes and present characterization of control policies by discrepancy functions. In contrast to the standard models on control of semi-Markov processes we assume that the decision maker has complete information on random times spent in each state (only on mean time not only on the mean time spent in each state) along with complete information on the progress of the controlled. On analyzing the current state of the process can decide if some (costly) intervention changing the current state of the process is suitable.

We begin our analysis with so called unichain models, i.e. when the underlying Markov chain contains a single class of recurrent states and hence the resulting average reward per unit time is independent of the starting state. Our analysis can be easily extended to a more general multichain model where average reward per time depends on the starting state and the state space can be partitioned on classes with the same average reward.

2.1 Unichain models

To begin with we make

Assumption 1. There exists state $i_0 \in \mathcal{I}$ that is accessible from any state $i \in \mathcal{I}$ for every $f \in \mathcal{F}$.

Obviously, if Assumption 1 holds, then the resulting transition probability matrix $P(f)$ is *unichain* for every $f \in \mathcal{F}$ (i.e. $P(f)$ has no two disjoint closed sets).

At first we focus attention on the embedded Markov chains and slightly extend some results reported in [8]. To this end, on introducing for arbitrary $g, w_j \in \mathbb{R}$ $(i, j \in \mathcal{I})$ and decision $f \in \mathcal{F}$, the discrepancy functions

$$\varphi_{i,j}^{\mathrm{c}}(w^{\mathrm{c}}, g^{\mathrm{c}}, f) := d_i(f_i) \cdot r(i) + r_{ij} - w_i^{\mathrm{c}} + w_j^{\mathrm{c}} - g^{\mathrm{c}}, \tag{1}$$

$$\varphi_{i,j}^{\mathrm{t}}(w^{\mathrm{t}}, g^{\mathrm{t}}, f) := d_i(f_i) - w_i^{\mathrm{t}} + w_j^{\mathrm{t}} - g^{\mathrm{t}} \tag{2}$$

for the random reward obtained, resp. time elapsed, up to the nth transition we have

$$\xi_n = ng^{\mathrm{c}} + w^{\mathrm{c}}_{X_0} - w^{\mathrm{c}}_{X_n} + \sum_{k=0}^{n-1} \varphi^{\mathrm{c}}_{X_k,X_{k+1}}(w^{\mathrm{c}}, g^{\mathrm{c}}, f), \tag{3}$$

$$\tau_n = ng^{\mathrm{t}} + w^{\mathrm{t}}_{X_0} - w^{\mathrm{t}}_{X_n} + \sum_{k=0}^{n-1} \varphi^{\mathrm{t}}_{X_k,X_{k+1}}(w^{\mathrm{t}}, g^{\mathrm{t}}, f). \tag{4}$$

Hence by (3), (5) for the expectation of ξ_n, $\mathsf{E}^{\,\pi}_i \xi_n =: v^{\pi}_i(n)$, resp. of τ_n, with $\mathsf{E}^{\,\pi}_i \tau_n =: t^{\pi}_i(n)$, we get

$$v^{\pi}_i(n) = ng^{\mathrm{c}} + w^{\mathrm{c}}_i + \mathsf{E}^{\,\pi}_i\{\sum_{k=0}^{n-1} \varphi^{\mathrm{c}}_{X_k,X_{k+1}}(w^{\mathrm{c}}, g^{\mathrm{c}}, f) - w^{\mathrm{c}}_{X_n}\}, \tag{5}$$

$$t^{\pi}_i(n) = ng^{\mathrm{t}} + w^{\mathrm{t}}_i + \mathsf{E}^{\,\pi}_i\{\sum_{k=0}^{n-1} \varphi^{\mathrm{t}}_{X_k,X_{k+1}}(w^{\mathrm{t}}, g^{\mathrm{t}}, f) - w^{\mathrm{t}}_{X_n}\}. \tag{6}$$

Now we show how to express average reward generated by the semi-Markov process $Y(t), t \geq 0$ in terms of the embedded Markov chain X_n. Considering policy $\pi \sim (f)$, let

$$\varphi^{\mathrm{c}}_i(w^{\mathrm{c}}, g^{\mathrm{c}}, f) \;:=\; \sum_{j\in\mathcal{I}} p_{ij}(f_i)\varphi^{\mathrm{c}}_{i,j}(w^{\mathrm{c}}, g^{\mathrm{c}}, f) = \sum_{j\in\mathcal{I}} p_{ij}(f_i)[d_i(f_i)\cdot r(i) + r_{ij} - w^{\mathrm{c}}_i + w^{\mathrm{c}}_j - g^{\mathrm{c}}], \tag{7}$$

$$\varphi^{\mathrm{t}}_i(w^{\mathrm{t}}, g^{\mathrm{t}}, f) \;:=\; \sum_{j\in\mathcal{I}} p_{ij}(f_i)\varphi^{\mathrm{t}}_{i,j}(w^{\mathrm{t}}, g^{\mathrm{t}}, f) = \sum_{j\in\mathcal{I}} p_{ij}(f_i)[d_i(f_i) - w^{\mathrm{t}}_i + w^{\mathrm{t}}_j - g^{\mathrm{t}}] \tag{8}$$

It is well-known from the dynamic programming literature (cf. e.g. [1, 3, 5, 6]) that for every $f \in \mathcal{F}$ and arbitrary transition costs $s_{ij}(f) = d_i(f_i)r(i) + r_{ij}, i, j, \in \mathcal{I}$, there exist numbers $g(f)$ and $w_i(f), i \in \mathcal{I}$ (unique up to additive constant) such that

$$w_i(f) + g(f) = d_i(f_i)r(i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij} + w_j(f)], \;\; (i \in \mathcal{I}), \;\; \text{i.e.} \tag{9}$$

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,\varphi^{\mathrm{c}}_{i,j}(w, g, f) = 0 \quad \text{where} \;\; \varphi^{\mathrm{c}}_{i,j}(w, g, f) := d_i(f_i)r(i) + r_{ij} - w_i(f) + w_j(f) - g(f).$$

In particular, for suitable selected $w^{\mathrm{c}}_j(f)$, resp. $w^{\mathrm{t}}_j(f)$, we have

$$v^{\pi}_i(n) = ng^{\mathrm{c}}(f) + w^{\mathrm{c}}_i(f) - \sum_{j\in\mathcal{I}} p_{ij}(f)\cdot w^{\mathrm{c}}_j(f), \quad \text{where} \tag{10}$$

$$w^{\mathrm{c}}_i(f) + g^{\mathrm{c}}(f) = d_i(f_i)\cdot r(i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij} + w^{\mathrm{c}}_j(f)],$$

Similarly, for suitable selected $w^{\mathrm{t}}_j(f)$, we have

$$t^{\pi}_i(n) = ng^{\mathrm{t}} + w^{\mathrm{t}}_i - \sum_{j\in\mathcal{I}} p_{ij}(f)\cdot w^{\mathrm{t}}_j(f), \quad \text{where} \tag{11}$$

$$w^{\mathrm{t}}_i(f)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} + g^{\mathrm{c}}(f) = d_i(f_i)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot w^{\mathrm{t}}_j(f)\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} \tag{12}$$

and by subtracting (12) from (11) we get

$$w_i(f) = \bar{r}_i(f) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)w_j(f) - d_i(f_i)g(f) \tag{13}$$

with

$$w_i(f) := w^{\mathrm{c}}_i(f) - w^{\mathrm{t}}_i(f)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)}, \quad g(f) := \frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)}, \quad \bar{r}_i(f) = d_i(f_i)\cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} p_{ij}(f_i)r_{ij}.$$

On introducing matrix notations

$$P(f) = [p_{ij}(f_i)], \quad D(f) = \text{diag}\,[d_i(f_i)], \ \text{(square matrices)}$$

$$\bar{r}(f) = [\bar{r}_i(f)], \quad w(f) = [w_i(f)], \quad \bar{g}(f) = [g(f)] \ \text{(column vectors)}$$

equation (11) can be written as

$$w(f) = \bar{r}(f) + P(f)w(f) - D(f)\bar{g}(f) \Rightarrow \bar{g}(f) = D^{-1}(f)\bar{r}(f) + [D^{-1}(f)P(f) - I] \cdot w(f)]. \tag{14}$$

Let

$$\tilde{r}(f) := D^{-1}(f)\bar{r}(f), \qquad \tilde{w}(f) := D^{-1}(f)w(f), \qquad \tilde{P}(f) := D^{-1}(f) \cdot P(f) \cdot D(f)$$

Then

$$\bar{g}(f) = \bar{r}(f) + [\bar{P}(f) - I] \cdot \bar{w}(f)]. \tag{15}$$

and for elements of $\tilde{r}(f),\ \tilde{w}(f),\ \tilde{P}(f)$ we have

$$\tilde{r}_i(f) = \bar{r}(i) + [d_i(f_i)]^{-1}r_{ij}, \quad \tilde{p}_{ij}(f_i) := p_{ij}(f_i)\frac{[d_j(f_j)]}{[d_i(f_i)]}, \quad \tilde{w}_i(f) := [d_i(f_i)]^{-1}w_i(f)$$

(observe that $\bar{g}(f)$ is a constant vector with elements $g(f) = \frac{g^c(f)}{g^t(f)}$).

In particular, let us consider continuous-time Markov decision chain with transition intensities $\mu_{ij}(f_i)$, where $\sum_{j\in\mathcal{I},j\neq i}\mu_{ij}(f_i) = -\mu_{ii}(f_i)$ and $\mu_i(f_i) = -\mu_{ii}(f_i)$ is the is the intensity of jumps from state i. Obviously, this is a very special case of semi-Markov processes with transition probabilities $p_{ij}(f) = \frac{\mu_{ij}(f_i)}{\mu_i(f_i)}$, and expected holding time $d_i(f_i) = \frac{1}{\mu_i(f_i)}$ in state i. Then on replacing in (14) transition probabilities and expected holding times by transition intensities for the average reward per unit of time of the considered continuous-time Markov process we conclude that

$$g(f) = r(i) + \sum_{j\neq i}\mu_{ij}(f_i)r_{ij} + \sum_j \mu_{ij}(f_i)w_j(f) \tag{16}$$

and policy $f \in \mathcal{F}$ appears in equation for average reward of a continuous time Markov reward chain (cf. e.g. [3]).

2.2 Multichain models

Considering transition probability matrix $P(f)$ for fixed $f \in F$, as well-known it is possible decompose the considered Markov chain such that

$$P(f) = \begin{bmatrix} P_{00}(f) & P_{01}(f) & P_{02}(f) & \dots & P_{0r}(f) \\ 0 & P_{11}(f) & 0 & \dots & 0 \\ 0 & 0 & P_{22}(f) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & \dots & P_{rr}(f) \end{bmatrix} \tag{17}$$

where existence of at least one recurrent state accessible from any state of $P_{ii}(f)$ (cf. Assumption 1) is fulfilled for diagonal submatrices $P_{ii}(f)$, $i = 1, \dots, r$, and $P_{00}(f)$ contains transient states that have access at least to two diagonal submatrices $P_{ii}(f)$, $i = 1, \dots, r = r(f)$.Observe that each diagonal block of $P(f)$ contains a single class of recurrent states possibly along with transient states accessible only to recurrent states of the same diagonal block.

In what follows we show that the decomposition depicted in (17) can be extended to the set of all matrices $P(f)$ with $f \in \mathcal{F}$. To this end, on recalling the notions of accessibility and communication for states of Markov chains, on considering stationary policies, say $f^{(1)}, f^{(2)}$, then for any $\kappa_i \in (0, 1)$ there exists $f \in \mathcal{F}$ such that for any $i, j \in \mathcal{I}$ it holds $p_{ij}(f) = \kappa_i p_{ij}(f^{(1)}) + (1 - \kappa_i)p_{ij}(f^{(2)})$. Hence any recurrent class of $P(f^{(1)})$ or $P(f^{(2)})$ must be contained in some recurrent class of $P(f)$.

Repeating this analysis for all admissible stationary policies we can conclude existence of (possibly randomized)

stationary policy, say $\tilde{f}$, such that

$$P(\tilde{f}) = \begin{bmatrix} P_{00}(\tilde{f}) & P_{01}(\tilde{f}) & P_{02}(\tilde{f}) & \dots & P_{0r}(\tilde{f}) \\ 0 & P_{11}(\tilde{f}) & 0 & \dots & 0 \\ 0 & 0 & P_{22}(\tilde{f}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & \dots & P_{rr}(\tilde{f}) \end{bmatrix} \tag{18}$$

with $r = \min r(\tilde{f})$ for any $f \in \mathcal{F}$.

Recall that Assumption 1 of Section 2.1 is fulfilled for each diagonal submatrix $P_{ii}(\tilde{f})$, $i = 1, \dots, r$ and $P_{00}(\tilde{f})$ contains transient states that have access at least to two diagonal submatrices $P_{ii}(\tilde{f})$, $i = 1, \dots, r = r(\tilde{f})$. Moreover observe that for $i = 1, \dots, r$ submatrix $P_{ii}(\tilde{f})$, contains recurrent class that is final, i.e. elements of this class have no access to states not belonging to this class, in contrast to elements of $P_{00}(\tilde{f})$. Employing the decomposition according to (18) we can employ policy iteration to find stationary policy, say $\hat{\pi} \sim \hat{f}$, yielding maximal average reward, say g_i^*, for elements of every diagonal submatrix $P_{ii}(\tilde{f})$. For details see [3].

Using this approach we finally arrive to the following decomposition

$$P(\hat{f}) = \begin{bmatrix} P_{00}(\hat{f}) & P_{01}(\hat{f}) & P_{02}(\hat{f}) & \dots & P_{0r}(\hat{f}) \\ 0 & P_{11}(\hat{f}) & 0 & \dots & 0 \\ 0 & 0 & P_{22}(\hat{f}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & \dots & P_{rr}(\hat{f}) \end{bmatrix} \tag{19}$$

with $r = \min r(\hat{f})$ for any $f \in \mathcal{F}$.

3 Improving Reward Optimality by Decision Maker Interventions

Stationary policy $\hat{\pi} \sim \hat{f}$ constructed in the previous section maximizes long-run average reward of the considered semi-Markov process. Since the decision maker has complete information on the current states in any time instant, and can change the action with respect the current state of the process, the following question can be raised:

Supposing that the decision maker considers that the current state of the process is suitable, is it suitable to change the current state by transfer the process to a more suitable state. If course, such a transfer is costly and the following question can be raised: Is it suitable for a given penalty cost to transfer the process to another state. To this end, we construct an improved stationary policy and compare long-run average reward of the original and improved policies.

Illustrative example.

Consider the controlled process with 6 states and suppose that equation (19) can be decomposed such that

$$P(\hat{f}) = \begin{bmatrix} p_{11}(\hat{f}) & p_{12}(\hat{f}) & p_{13}(\hat{f}) & p_{14}(\hat{f}) & p_{15}(\hat{f}) & p_{16}(\hat{f}) \\ p_{21}(\hat{f}) & p_{22}(\hat{f}) & p_{23}(\hat{f}) & p_{24}(\hat{f}) & p_{25}(\hat{f}) & p_{26}(\hat{f}) \\ 0 & 0 & p_{33}(\hat{f}) & p_{34}(\hat{f}) & 0 & 0 \\ 0 & 0 & p_{43}(\hat{f}) & p_{44}(\hat{f}) & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{55}(\hat{f}) & p_{56}(\hat{f}) \\ 0 & 0 & 0 & 0 & p_{56}(\hat{f}) & p_{66}(\hat{f}) \end{bmatrix} \tag{20}$$

Using the decomposition according to (17), optimal policy contains along with transient states also two (final) classes of recurrent states, as shown in the following display.

In particular,

$$P(\hat{f}) = \begin{bmatrix} P_{00}(\hat{f}) & P_{01}(\hat{f}) & P_{02}(\hat{f}) \\ 0 & P_{11}(\hat{f}) & 0 \\ 0 & 0 & P_{22}(\hat{f}) \end{bmatrix} \tag{21}$$

where

$$P_{00}(\hat{f}) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}, P_{01}(\hat{f}) = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{bmatrix}, P_{02}(\hat{f}) = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{bmatrix}, P_{11}(\hat{f}) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{6} & \frac{5}{6} \end{bmatrix}, P_{22}(\hat{f}) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

We assume that the submatrix $P_{11}(\hat{f})$, resp. submatrix $P_{22}(\hat{f})$, generates average reward $G_1(\hat{f})$, resp. $G_2(\hat{f})$. Hence if the process starts in state 1 or 2, the resulting average reward is a suitable linear combination of the values of average rewards $G_1(\hat{f})$ and $G_2(\hat{f})$.

Supposing that the considered process starts in state 1 or 2, if $G_1(\hat{f}) > G_2(\hat{f})$ it is possible by the decision maker's intervention to stop reaching states 5 and 6 by changing submatrices $P_{01}(\hat{f})$, $P_{01}(\hat{f})$ to

$$P_{01}(\hat{f}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad P_{02}(\hat{f}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

References

[1] Bertsekas, D. P. (2007). *Dynamic Programming and Optimal Control,* Volume 2, Third Edition. Belmont, Mass.: Athena Scientific.

[2] Gantmakher, F. R. (1959). *The Theory of Matrices.* London: Chelsea.

[3] Howard, R. A. (1960). *Dynamic Programming and Markov Processes.* Cambridge, Mass.: MIT Press.

[4] Howard, R. A. & Matheson, J. (1972). Risk-sensitive Markov decision processes, *Manag. Sci.*, 23 , 356–369.

[5] Puterman, M. L. (1994). *Markov Decision Processes – Discrete Stochastic Dynamic Programming*. New York: Wiley.

[6] Ross, S. M. (1983). *Introduction to Stochastic Dynamic Programming.* New York: Academic Press.

[7] Sladký, K. (1973). Necessary and sufficient optimality conditions for average reward of controlled Markov chains, *Kybernetika*, 9, 124–137.

[8] Sladký, K. (2005). On mean reward variance in semi-Markov processes, *Math. Methods Oper. Res.*, 62, 387–39

[9] Sladký, K. (2012). Risk-sensitive and average optimality in Markov decision processes. In J. Ramík & D. Stavárek (Eds.), *Proc. 30th Internat. Conference Mathematical Methods in Economics 2012, Part II* (pp. 799–804). Karviná: Silesian University, School of Business Administrations.

[10] van Dijk, N. M. & Sladký, K. (2006). On the total reward variance for continuous-time Markov reward chains, *J. Appl. Probab.*, 43, 1044–1052.

Priority Single-Server Queuing System with Optional Second Server Activated upon Request – Simulation Study

Rostislav Stryk[1], Abate Getaw Sewagegn[2], Petr Jaluvka[3], Michal Dorda[4]

Abstract. Queueing systems with priorities can often be met in practice - e.g. in healthcare or transport. These systems are specific in that there are at least two classes of customers, while priorities are defined for individual classes of customers, according to which customers are served - the priority can be non-preemptive (higher priority customers do not interrupt the service of a lower priority customer) or preemptive-resume (higher priority customers interrupt the service of a lower priority customer). In this paper, we focus on a queueing system with three classes of customers, where the first class has the highest priority and the last third class has the lowest priority. By default, individual customers entering the system are served by single server. However, if a situation arises where the queue length of non-priority customers exceeds a certain defined threshold, it is possible to activate a second parallel server, which helps with serving the lower priority customers. The article describes the simulation model of this queueing system and, based on the experiments carried out, recommendations for the activation of the second server are defined.

Keywords: priority queue, optional server, simulation, colored Petri nets, air traffic control

JEL Classification: C44
AMS Classification: 60K25

1 Introduction and Our Motivation to Develop the Model

Priority single-server queueing models can be used for modelling a lot of practical systems for example in transport or informatics. For priority queueing systems it is typical that such systems serves several classes of customers which priorities differ. In general we can say the priority can be non-preemptive (higher priority customers do not interrupt the service of a lower priority customer) or preemptive-resume (higher priority customers interrupt the service of a lower priority customer) – see more information for example in [1] or [8]. In article [2] a priority queuing model application in an operating system is shown. The researcher indicated that it is possible to model a single-server priority queuing system with several classes of customers. Source [4] presented a parallel priority queuing system, and achieved results indicated that the system reduced the idle period of the server. Priority queuing systems are applicable in different areas, like hospitals; see for example paper [3] which discussed priority queuing models for hospital intensive care units.

The idea of a second optional server activated based on customer request has been investigated by several researchers. Source [5] discussed batch arrival priority queuing with a second optional service and server breakdown. A priority queuing model is investigated in paper [7]. Researchers applied Petri Net and concluded that the tool is convenient, efficient, and effective for modeling priority queuing.

Because of defined priorities it may happen that the lower priority customers must wait too long to be serviced, especially in cases when the occupation rate (usually denoted as ρ) of the highest priority customers tends to 1. To reduce waiting times of the lower priority customers in such cases an optional (secondary) server may be activated to help the primary server to serve the customers waiting in the queue – the additional server can serve all classes of customers or customers of selected classes only. In queueing theory, such optional server is usually activated when the number of customers waiting in the queue reaches a pre-defined value – this is called a threshold. The

[1] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, rostislav.stryk@vsb.cz
[2] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, abate.getaw.sewagegn.st@vsb.cz
[3] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, petr.jaluvka@vsb.cz
[4] VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, michal.dorda@vsb.cz

threshold can be related to all the customers waiting for service or only to selected classes of customers. After reaching this threshold the optional server can be activated immediately (this is usually true in cases when the optional server is found in the system permanently) or it takes a certain time to activate the optional server lasts – the time needed can be deterministic or stochastic – this assumption is usually done when the optional server is not found in the system permanently. The optional server is usually deactivated when the number of customers waiting in the queue reaches zero.

Our main idea to develop the model is based on modelling air traffic control processes which can be modelled as a queueing system with three customer classes differing in the priorities. Class A, with the higher priority, includes control service provision of safety instructions, information and clearances preventing aircraft collisions and maintaining expeditions flow of air traffic. Class B, with the medium priority, includes provision of clearance delivery service to IFR (Instrument Flight Rules) flights and class C, with the lowest priority level, includes provision of clearance delivery service to VFR (Visual Flight Rules) flights.

The paper describes how such priority queueing system can be modelled using colored Petri nets and results of experiments done with the model. The aim of the experiments was to find the moment, when the mean waiting time of the IFR flights requests for the ATC clearance delivery reaches a specific time described as a maximum acceptable mean waiting time. In that moment the new separate working position for the ATC clearance delivery service must be activated to provide the ATC clearance delivery service. The maximum acceptable mean waiting time for the ATC clearance delivery service for IFR and VFR flights was set to 4 minutes for the experiment purposes.

2 General Queueing Model

Let us consider a queueing system which serves three classes of customers with different priorities – let us denote the classes as A, B and C. We assume that the class A customers have the priority over class B and C customers (they have the highest priority). The class B customers have the priority over the class C customers (they have the medium priority). The class C customers have the lowest priority. Let us consider the priority is non-preemptive – that means a higher priority customer does not interrupt the service of a lower priority customer when entering the system.

Let us consider that interarrival times of the customers follow an exponential distribution of probability with the mean inter-arrival times $\frac{1}{\lambda_A}$, $\frac{1}{\lambda_B}$ and $\frac{1}{\lambda_C}$. The customers waiting in the queue are served by a single primary server. Let the service times be exponentially distributed with the mean service times $\frac{1}{\mu_A}$, $\frac{1}{\mu_B}$ and $\frac{1}{\mu_C}$.. However, when the number of the class B and C customers reach a certain threshold *a* then an additional server is activated – the server helps to serve the customers of classes B and C. Let us assume that the additional server is activated (is prepared to serve customers) immediately upon request (there is no delay). The server is deactivated when the number of class B and C customers waiting in the queue reaches 0.

3 Simulation Model

This paper tries to model the process of priority customer service with second optional server. To model this priority queueing model CPN tools in version 4.0.1 was applied; CPN Tools represent a tool which has been designed for editing, simulating, and analyzing Colored Petri nets – for more information see https://cpntools.org/. More information about colored Petri nets can be seen for example in [6] or [9].

Each CPN model is graphically presented by a directed bipartite graph in which vertices represent so-called places and transitions and arcs connect individual places with transitions and vice versa – see for example [8]. In general, places of Colored Petri nets model individual states of the system and transitions model changes of the individual states. It is possible to use "basic" directed arcs, testing arcs (they are depicted as two-directional arcs) and inhibitor arcs (they are depicted as directed arcs but with a circle instead of arrow). Individual states of the system are defined by so-called marking of the net which is given by distribution of tokens of individual colors over the individual places of the net. The graphical part of the CPN model is supplemented by so-called inscription which comprise definition of arc inscriptions, initial marking of individual places, transition priorities, color sets, variables and functions used in the model etc.

The petri net presented in Figure 1 is composed of 18 places and 11 transitions which are connected with arcs. Tokens applied to represent customers, server and the flow of tokens indicate the movement of customers. To create the model it was necessary to define some color sets, variables and functions:

- a color set declared as "*colset UNIT = unit;*" enables to work with basic tokens denoted ().
- a color set defined as "*colset UNITtm = unit timed;*" is derived from the previous color set, the difference is that tokens () are equipped with a time stamp.
- a color set declared as "*colset INTINF = intinf;*" enables to work with tokens which color is expressed as a non-negative integer value.
- a color set defined as "*colset INTINFlist = list INTINF;*" enables using lists containing non-negative integer values in the model.
- a variable declared as "*var customers :INTINFlist;*" is used to handle with lists from the precious inscription.
- a variable defined as "*var customer: INTINF;*" enables to assign tokens a color expressed as a non-negative integer value.
- a color set declared as "*colset p = with x | y timed;*" is used to model the servers – in the model we assume two servers – a primary one which is modelled by a token of color *x* and a secondary one modelled by a token of color *y*. The color set is defined as timed to model service times.
- a variable defined as "*var server: p;*" can take values from the color set *p* and therefore can be used to assigning a corresponding server to the service process.
- a function declared as "*fun ET(EX) = round (exponential(1.0/EX));*" generates values coming from the exponential distribution of probability defined by the mean value *EX*.

The simulation model represented by Figure 1 can be divided into 4 parts which are highlighted red, blue, green and purple. The elements highlighted red model input of the highest priority customers (denoted as customers of class A in the model), their waiting for service and their service. The same meaning have the elements highlighted blue and green, but they are used for modelling the same processes of lower priority customers – customers of class B (blue color) and customers of class C (green color). The remaining elements highlighted purple model sources (servers) of the model.

Because the parts modelling the processes for the individual classes are similar, let us describe briefly the meaning of the individual elements for the customers of class A. The place "*custo_A rate*" together with the transition "*Customer A*" including the corresponding arcs model the input of the class A customers to the system, in the model it is assumed that the inter-arrival times are exponentially distributed (the Poisson input stream). The places "*Queue A*" and "*custo_A waiting for service*" model waiting the class A customers in the queue to be serviced. The place "*Queue A*" contains a token from a color set "*INTINFlist*" – the token represent a list of customers waiting in the queue – each customer waiting in the queue is represented by a non-negative integer value in this list; the value correspond to the simulation time when the customer has entered the queue. Each time a new customer enters the queue the value is added at the end of the list and each time a customer exits the queue the first value in the list is removed – that correspond to the FIFO service discipline for the class A customers. To be able to estimate the mean number of the class A customers waiting in the queue the place auxiliary "*custo_A waiting for service*" is used – the marking of the place represent the number of the class A customers waiting in the queue.

By firing the transition "*custo_A service start*" class A customers exit the queue and their service is started. Because of the highest priority of the class A customers the priority of the transition is set to "*P-HIGH*". Please note, that the corresponding transitions for the lower priority customers have their priorities set to "*P-NORMAL*" (this is a standard priority in CPN Tools so it is not depicted in Figure 1) and "*P-LOW*". That means if all these transitions would be enabled at the same time, the transition with the highest priority is fired.

The place "*custo_A service*" model the service of the class A customers, the service times are considered to be exponentially distributed – the service times are modelled by increasing the time stamp of the token entering the place via the arc inscription connecting the transition "*custo_A service start*" with the place. The service is terminated when the actual simulation time equals to the time stamp of the token found in the place "*custo_A service*". And finally, the auxiliary place "*Arrived custo_A*" is used to monitor the number of the class A customers in the system.

As stated earlier, the parts of the model highlighted blue and green fulfill the same function as the red highlighted part. The differences lie in the mean inter-arrival and service times and the priorities of transitions "*custo_B service strt*" and "*custo_C service strt*". The last difference is that the class A customers are served by the primary server modelled by the token of color "*x*" only whilst the class B and C customers can be served by any of the servers.

The last part of the simulation model highlighted purple models the servers of the queuing model. The place "*Free server*" contains tokens that represent the idle servers found in the system – the initial marking of the place is defined as "*1`x*" because the secondary server enters the system when the threshold is reached. A token of color "*y*" found in the place "*Sec. server out of system*" represent the additional (secondary) server when this server is

not present in the system. By firing the transition "*Activate sec. server*" the additional server is activated; the transition is enabled when the number of the class B and C customers waiting in the queue, which is determined on the basis of marking of the auxiliary place "*Non-priority queue*", has reached the threshold value *a*. On the other hand, the additional served is deactivated (it leaves the system) by firing the transition "*Deactivate sec. server*" which is enabled when no class B and C customers are waiting to be serviced.

After we created the Petri net model, some monitoring function to estimate desired outputs – performance measures – of the model were defined. The main performance measures are the following:

- Mean numbers of the customers of each class waiting in the queue – EL_A, EL_B, and EL_C; the performance measures can be estimated by monitoring markings of places "*custo_A waiting for service*", "*custo_B waiting for service*", "*custo_C waiting for service*".
- Mean numbers of the customers of each class in service – ES_A, ES_B, and ES_C; the performance measures can be estimated by monitoring markings of places "*custo_A service*", "*custo_B service*", "*custo_C service*".
- Mean waiting times of the customers of each class – EW_A, EW_B, and EW_C; ; the performance measures can be estimated by means of data collection monitoring functions associated with the transitions "*custo_A service start*", "*custo_B service start*", "*custo_C service start*".
- Probability that the additional server is out of the system – P_{out}; the performance measure is estimated via monitoring markings of the place "*Sec. server out of system*". The value $\kappa_{additional} = 1 - P_{out}$ then represents the mean utilization of the additional server because the additional server is always busy when active.
- Mean idleness of the primary server – *IDL*; the performance measure is estimated by means of monitoring markings of the place "*Free server*". The value $\kappa_{primary} = 1 - IDL$ then represents the mean utilization of the primary server.

In this Petri net model we applied a breakpoint monitoring function to terminate each simulation run at simulation time 2952000 seconds (30 days). Using the auxiliary text "*CPN'Replications.nreplications 10*" allows to run the simulation model 10 times.

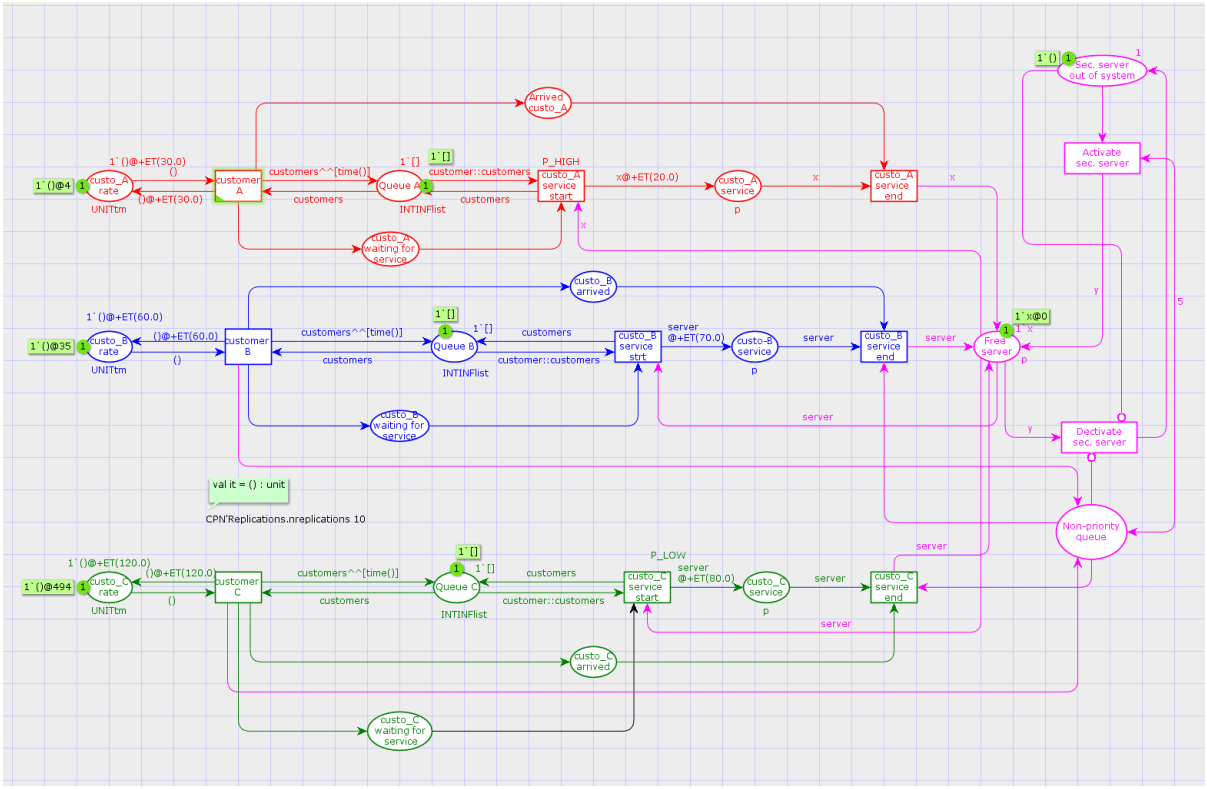


Figure 1 Priority queueing model with optional server created using CPN tools

4 Experiments and Their Results

To demonstrate the impact of the additional server on the performance measures of the modelled queueing system we run some experiments – a summary of the model parameters related to the individual customer classes is presented in Table 1. Please note that it is assumed that the inter-arrival and service times are exponentially distributed.

Customer class	The arrival rate λ [customers per an hour]	The mean inter-arrival time $\frac{1}{\lambda}$ [seconds per a customer]	The service rate μ [customers per an hour]	The mean service time $\frac{1}{\mu}$ [seconds per a customer]
A	120.00	30.00	180.00	20.00
B	10.00	360.00	80.00	45.00
C	10.00	360.00	60.00	60.00

Table 1 Summary of the model parameters

At first we can run an experiment in which the additional server cannot be activated. This can be done for example by removing the initial marking of the place "*Sec. server out of system*". Results of this experiment are presented in Table 2 and Table 3 in the first row (for the threshold ∞). As we can see in Table 2, the mean waiting time for the class B customers is about 7 minutes and for the class C customers about 57 minutes. The mean waiting times for these customers are high because of the total occupation rate ρ for which it holds:

$$\rho = \rho_A + \rho_B + \rho_C = \frac{\lambda_A}{\mu_A} + \frac{\lambda_B}{\mu_B} + \frac{\lambda_C}{\mu_C} = \frac{120}{180} + \frac{10}{80} + \frac{10}{60} \doteq 0.96.$$

The total occupation rate is close 1 – this is the reason why the mean waiting times of the class B and C customers are so high, the system is close to its capacity. In relation to the possible application of the model we can say that such mean waiting times are not acceptable.

To reduce the mean waiting times of the lower priority customers we can activate the additional server. We run several experiments differing in the threshold which was considered to be from 1 up to 10 – see results of the experiments in Table 2 and Table 3. Please note that the presented results represent point estimations of the monitored performance measures.

Threshold	EL_A [-]	EL_B [-]	EL_C [-]	EW_A [s]	EW_B [s]	EW_C [s]
∞	2.894170	1.177603	9.580548	86.730119	422.396469	3435.618120
1	1.666295	0.020866	0.027520	50.035036	7.473406	9.917719
2	2.077800	0.111106	0.145078	62.324038	40.205697	52.146225
3	2.275739	0.221929	0.304877	68.242201	79.632576	109.356722
4	2.330481	0.318390	0.471683	69.980092	114.571744	169.781401
5	2.420545	0.409281	0.654750	72.679539	145.852604	235.386085
6	2.474298	0.490024	0.839530	74.191984	174.967610	302.410701
7	2.500887	0.540632	1.035235	75.047260	195.837219	373.187031
8	2.559997	0.604514	1.222581	76.881626	217.379595	438.424601
9	2.571804	0.659424	1.446040	77.242931	237.048405	519.509087
10	2.605583	0.708626	1.652954	78.253190	256.527501	594.430156

Table 2 Results of the simulation experiments – part 1

The experimental results show the threshold when the optional server should be activated if we want not to exceed given mean wating times for the class B and C customers. It was stated in the introduction that the maximal acceptable mean wating time is 4 minutes (240 seconds). This condition is met for the thresholds up to 5. As can be further seen in the results, for the threshold 1 the mean waiting times of the class B and C customers are even lower than the mean waiting time of the class A customers which have the highest priority – this is caused by the fact that with this threshold the additional server is activated each time a class B or C customer enters the queue. In addition this threshold brings the lowest utilization of the primary server which is about 68.7 percent because of the low threshold most class B and C customers are served by the additional server. On the other hand, for the threshold 5 the mean waiting times of the class B and C customers are lower than the predefined limit whilst the utilization of the primary server is about 86.8 percent.

Threshold	ES_A [-]	ES_B [-]	ES_C [-]	P_{out} [-]	*IDL* [-]
∞	0.666944	0.125820	0.166441	1.000000	0.040794
1	0.664997	0.124806	0.166571	0.730900	0.312727
2	0.667011	0.123837	0.167111	0.829198	0.212843
3	0.667929	0.126166	0.167076	0.869327	0.169501
4	0.665477	0.124801	0.166475	0.894359	0.148888
5	0.666973	0.125803	0.167052	0.908543	0.131628
6	0.667080	0.125872	0.166335	0.920796	0.119917
7	0.665312	0.124557	0.166007	0.932254	0.111871
8	0.666084	0.125342	0.167180	0.937426	0.103968
9	0.666029	0.124750	0.166744	0.944873	0.097604
10	0.665395	0.123667	0.166794	0.952196	0.091948

Table 3 Results of the simulation experiments – part 3

5 Conclusions

In the article we presented the simulation model of the queueing model which serve three classes of customers differing in their priorities. To reduce the waiting times of the lower priority customers the additional (secondary) server is activated. The presented model was created to model air traffic control processes in order to estimate the threshold for activating the additional server.

Based on the experimental results we can define such threshold for which the mean waiting times of the lower classes of customers do not exceed pre-defined values. Please note that the experiments presented in the article are only preliminary because we have no precise data on arrival and service rates – the values used in the article are only their estimations. In our future research we want to collect all necessary input data to use the presented model for defining the threshold, when the additional server should be activated, for air traffic control processes at Ostrava Mosnov airport.

Acknowledgements

The paper was supported by the internal project of the Faculty of Mechanical Engineering, VSB – Technical University of Ostrava, SP2023/087 Applied research, experimental development and innovation in transport and logistics.

References

[1] Adan, I. & Resing, J. (2002). *Queueing Theory*. Eindhoven, Netherlands: Eindhoven Univ. Technol.

[2] Derbala, A. (2005). Priority queuing in an operating system. *Computers & operations research*, 32(2), 229-238.

[3] Hagen, M. S., Jopling, J. K., Buchman, T. G. & Lee, E. K. (2013). Priority queuing models for hospital intensive care units and impacts to severe case patients. In *AMIA annual symposium proceedings* (Vol. 2013, p. 841). American Medical Informatics Association.

[4] Haghighi, A. M. & Mishev, D. P. (2006). A parallel priority queueing system with finite buffers. *Journal of Parallel and Distributed Computing*, 66(3), 379-392.

[5] Jain, M. & Jain, A. (2014). Batch arrival priority queueing model with second optional service and server breakdown. *International Journal of Operations Research*, 11(4), 112-130.

[6] Jensen, K. & Kristensen, L. M. (2009). *Coloured Petri nets: modelling and validation of concurrent systems*. Springer Science & Business Media.

[7] Strzęciwilk, D. & Zuberk, W. M. (2019). Modeling and performance analysis of priority queuing systems. In *Software Engineering and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference 2018*, Volume 1 7 (pp. 302-310). Springer International Publishing.

[8] Thomopoulos, N. T. (2012). *Fundamentals of queuing systems: statistical methods for analyzing queuing models*. Springer Science & Business Media.

[9] Zaitsev, D. A. & Shmeleva, T. R. (2006). Simulating telecommunication systems with CPN Tools. *Students' Book.–Odessa: ONAT (nd*), 68.

Estimation of the Elasticity of Input Substitution in European Regions

Karol Szomolányi[1], Martin Lukáčik[2], Adriana Lukáčiková[3]

Abstract. The paper estimates the elasticity of input substitution in different European regions. European regional data is used for estimation. The chosen procedure focuses on estimating the econometric specification of labor demand. A frequency filter filters the time series in the econometric specification to adjust the relationship from other systematic economic processes. The short-term estimate is obtained by differentiating the variables of specification. The elasticity of substitution is estimated at the regional level as well as at the aggregate level. The research paper also estimates the aggregate elasticity of substitution according to regional growth rate and regional development level measured by real PPP GDP per person. The estimated elasticity of substitution varies in different countries, but in most cases, it is less than one, contrary to the Cobb-Douglas production function form. In aggregate, the estimate of the elasticity of substitution is about 0.6. A region's economic growth and development level have no impact on the estimated value of the elasticity of substitution. This result suggests that the European regions had a similar steady state during the studied period.

Keywords: elasticity of substitution, European regions, frequency filter, labor demand

JEL Classification: C32, E23
AMS Classification: 91B38, 91B84

1 Introduction

The value of the elasticity of substitution between labor and capital is a crucial assumption in understanding the secular decline in the labor share of income. For example, in DSGE models, its value affects the effect of price movements—including the traditional channel of monetary policy [2]. Moreover, it is a significant coefficient in the growth theory [3].

This contribution has the ambition to follow up Jones' [4] theoretical finding that the value of the elasticity of substitution of inputs differs from a local (regional) and global point of view. The paper aims to compare regional and aggregate estimates of the elasticity of substitution.

We highlight two of the many empirical approaches to estimating the elasticity of substitution. First, Klump et al. [5] estimated a system of 3 nonlinear equations; production function, capital, and labor demands. Second, Chirinko and Mallick [2] estimated the capital demand equation with data series modified by the low-pass filter to abstract them from the business cycles and the short-term effects driven by different underlying processes. This paper adopts the second one to estimate the elasticity of substitution using European regional data and labor demand econometric specification.

According to the approach of Chirinko and Mallick [2], the time series in the econometric specification are filtered by a frequency filter. The short-term estimate is obtained by differentiating the variables of specification. In addition, the frequency filter adjusts the relationship from other systematic economic processes.

Labor can be measured in two ways, by the total number of workers or the total number of hours worked. By both approaches, the elasticity of substitution is estimated on a regional and aggregate level. Estimates are realized using both dataset versions, the number of workers and hours worked. Moreover, using the frequency-filter panel data approach, the paper estimates the aggregate elasticity of substitution according to regional growth rate and regional development level measured by real PPS GDP per person.

[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.

The elasticity of substitution between inputs characterizes the relation between the input/output price ratio and the input/output ratio. Denoting the inputs capital and labor as K and L, respectively, their prices as w and r, respectively, and output as Y, its price P, relation (1) defines the elasticity. It expresses the negative percentual change of the input/output if the input/output price changes by 1%.

$$\sigma = -\frac{\partial\left(\frac{K}{L}\right)}{\partial\left(\frac{r}{w}\right)} \cdot \frac{\left(\frac{r}{w}\right)}{\left(\frac{K}{L}\right)} = -\frac{\partial\left(\frac{K}{Y}\right)}{\partial\left(\frac{r}{P}\right)} \cdot \frac{\left(\frac{r}{P}\right)}{\left(\frac{K}{Y}\right)} = -\frac{\partial\left(\frac{L}{Y}\right)}{\partial\left(\frac{w}{P}\right)} \cdot \frac{\left(\frac{w}{P}\right)}{\left(\frac{L}{Y}\right)} \tag{1}$$

Combining both relations, we can gain that elasticity of substitution also characterizes the relationship between the ratio of input prices and the input ratio (1). Notably, it expresses the negative percentual change of the input ratio K/L if the ratio of input prices w/r changes by 1 %.

Considering that, according to the first-order optimality condition of a firm maximizing its profit, the input/output price ratio equals the marginal product of the input, we can derive the production function in the form (2). One can find examples of the derivations in [5] or [3].

$$Y = Y_0 \left[\pi_0 \left(\frac{K}{K_0} \right)^{\frac{\sigma-1}{\sigma}} + (1-\pi_0) \left(\frac{L}{L_0} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

Deriving the constant elasticity of the substitution production function form requires using actual observed values of the variables Y_0, K_0, and L_0, including the distribution coefficients π_0, $1-\pi_0$. Therefore, we call the set of these variables the point of normalization, and in the production function form (2), it is marked with the subscript 0.

Adopting the graphs from the book of La Grandville [3], production function isoquants can present the elasticity of substitution (Figure 1) and the importance of normalization. Realizing that the slope of the isoquant corresponds to the input marginal product ratio (marginal rate of technological transformation) and the slope of the line connecting the coordinate system origin and the isoquant corresponds to the input ratio. The relation between the two slopes characterizes the elasticity of input substitution.

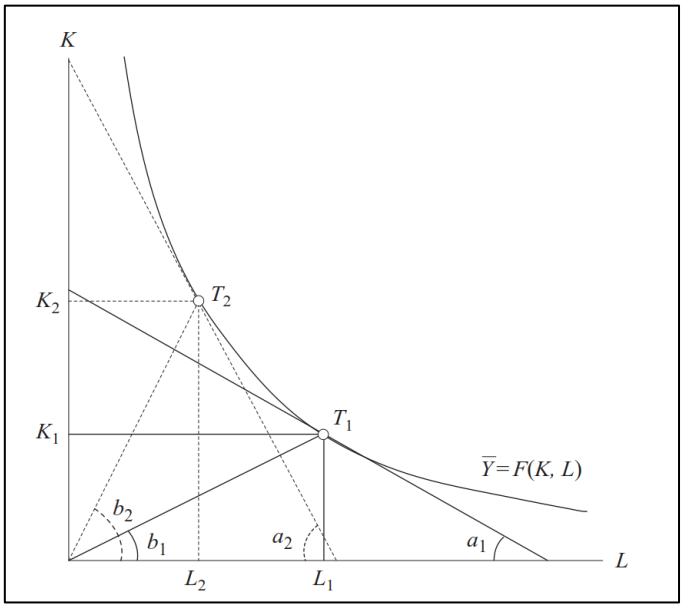


Figure 1 The elasticity of substitution, σ
Source: La Grandville [3]

Figure 2 highlights the importance of production function normalization. For a particular point of normalization, there is a map of isoquants corresponding to any elasticity of substitution value.

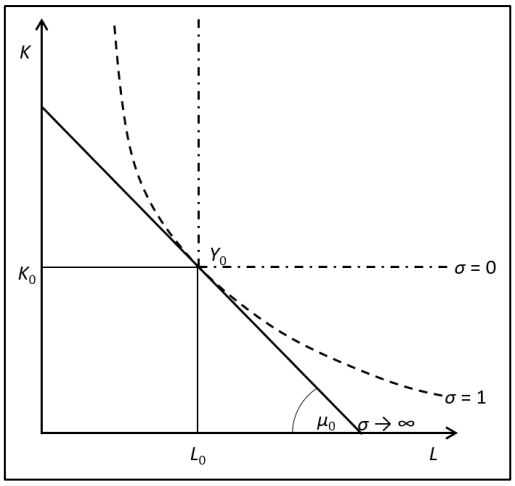


Figure 2 Map of isoquants
Source: La Grandville [3]

2 Methods and Data

The econometric specification of the relation (2) is in the form:

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_t + u_{it} \tag{3}$$

where y corresponds to the labor/output ratio, x corresponds to its relative prices (labor price/output price ratio), v_t corresponds to the trend term, and u corresponds to the stochastic term. Finally, both variables are measured by their natural logarithms to fit the relation (3).

To eliminate the effects of different underlying economic processes, ensuring the exogeneity of the price ratio x_t, both variables are measured by the proper long-run values. Chirinko and Mallick [2] argue that these values can be reached using a Low-Pass filter. The elasticity of substitution is the negative value of the β_1 coefficient.

The study uses NUTS 2 regional data of GDP at constant prices, GDP at current prices, total employment, total hours worked, and compensation of employees obtained from Eurostat. Consequently, the data range differs from time series to time series.

The input quantity data series is measured as total employment or hours worked, while the output quantity is measured as GDP at constant prices. The input price series can be computed as the compensation of employees/input quantity ratio. Output price series is GDP at current prices/GDP at constant prices ratio.

We used the Baxter and King [1] frequency filter to filter different underlying processes. The filter parameters correspond to the assumptions of the duration of the economic cycle of 2 to 8 years, and three leads and lags are considered for calculating the moving average. These settings are commonly recommended in the literature ([1], [2]).

We estimated the elasticity of input substitution by the least squares method of using single-equation econometric models for all regions and for both methods of measuring work, for which we had at least 20 observations of both time series available. Then using panel data, we estimated the aggregate elasticity of substitution for the whole European economy. Finally, we also used the panel data to estimate the elasticity of input substitution for different regional economic development and growth rate samples.

3 Results

Estimating the single-equation models, we state that the elasticity of input substitution is less than 1 in most cases. The average local elasticity of substitution is about 0.5. Tables 1, 2, and 3 show results for randomly chosen countries and their regions: Finland, Italy, and Portugal. Panel data estimations follow. First, using the total employment as the labor quantity, the econometric equation form is:

$$y_t = -\ 0.0067 - 0.6149 x_t + e_t \qquad R^2 = 0.6236$$
$$(s.e.)\ \ (0.0009)\ (0.0203)$$

If the total hours worked series are used, the equation is in the form:

$$y_t = -\ 0.0074 - 0.6185 x_t + e_t \qquad R^2 = 0.6502$$
$$(s.e.)\ \ (0.0008)\ (0.0221)$$

	employment		**hours**	
region	**sigma**	**std. error**	**sigma**	**std. error**
Länsi-Suomi	0.8409***	0.1905	0.8093***	0.1838
Helsinki-Uusimaa	1.0364***	0.0820	1.0089***	0.0728
Etelä-Suomi	0.9064***	0.1009	0.8672***	0.1113
Pohjois- ja Itä-Suomi	0.8929***	0.1039	0.8521***	0.1135
Åland	0.6319***	0.1944	0.9017***	0.2081

Table 1 The estimated results for the regions of Finland

	employment		**hours**	
region	**sigma**	**std. error**	**sigma**	**std. error**
Piemonte	1.5286***	0.2954	1.1583***	0.3474
Valle d'Aosta	0.9213***	0.1200	0.9651***	0.1355
Liguria	1.0969***	0.2159	0.8794***	0.1753
Lombardia	0.9151	0.4691	-0.4308	0.3675
Abruzzo	0.6794***	0.1876	0.5356***	0.1896
Molise	1.0765***	0.0932	1.0425***	0.0968
Campania	0.6432**	0.2535	0.3610	0.2821
Puglia	0.8397***	0.2979	0.6347**	0.2891
Basilicata	0.8136***	0.1368	0.8163***	0.1421
Calabria	0.9803***	0.1605	0.9807***	0.1895
Sicilia	0.4484	0.2591	0.2695	0.1943
Sardegna	0.8754***	0.1060	0.6940***	0.1726
Bolzano/Bozen	0.9593***	0.1258	0.9040***	0.1555
Trento	0.9009***	0.2109	0.5625**	0.2400
Veneto	0.7858**	0.3730	0.1721	0.4521
Friuli-Venezia Giulia	1.0534***	0.2104	0.5521**	0.2537
Emilia-Romagna	1.1628***	0.2574	0.3066	0.3638
Toscana	1.3015***	0.2702	0.4563	0.4044
Umbria	0.8767***	0.2698	0.4941	0.3689
Marche	0.8631***	0.1921	0.5718***	0.1863
Lazio	1.0002***	0.1858	0.7882***	0.2151

Table 2 The estimated results for the regions of Italy

	employment		**hours**	
region	**sigma**	**std. error**	**sigma**	**std. error**
Norte	0.4575***	0.0949	0.4295***	0.0851
Algarve	0.8133***	0.1409	0.7423***	0.1402
Centro (PT)	0.6178***	0.0832	0.6072***	0.0892
Área Metropolitana de Lisboa	0.4017***	0.1327	0.3850***	0.1149
Alentejo	0.8971***	0.0954	0.8586***	0.1074

Região Autónoma dos Açores	0.4851	0.3141	0.4942**	0.2091
Região Autónoma da Madeira	0.3940	0.3837	0.4758**	0.2282

Table 3 The estimated results for the regions of Portugal

Finally, estimates of the elasticity of input substitution for different regional economic development and growth rate samples were not statistically different.

4 Conclusion

This study confirmed some empirical findings from the literature. First, the elasticity of substitution is positive, less than 1. Klump et al. [5] and Chirinko and Mallick [2] support such a result with their empirical evidence and theoretical findings in other papers.

The paper's motivation was to confirm Jones' [4] theoretical supposition that the elasticity of input substitution is less locally than globally. The average regional estimation of the elasticity of input substitution is about 0.5, while the panel data global estimation is about 0.6. We estimate the global elasticity of input substitution to be only slightly higher than the local.

Following the growth theoretical suppositions elaborated in detail by La Grandville [3], we tried to find an empirical connection between the elasticity of input substitution and economic growth and development. However, we rejected the hypothesis that the elasticity of substitution is different in different regional economic development and growth rate samples.

Acknowledgements

The Grant Agency of Slovak Republic - VEGA supports this paper by grant no. 1/0211/21 "Econometric Analysis of Macroeconomic Impacts of Pandemics in the World with Emphasis on the Development of EU Economies and Especially the Slovak Economy" and by grant no. 1/0047/23 "The importance of spatial spillover effects in the context of the EU's greener and carbon-free Europe priority."

References

[1] Baxter, M. & King, R. G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics*, *81*, 575–593.

[2] Chirinko, R. S. & Mallick, D. (2017). The Substitution Elasticity, Factor Shares, and the Low-Frequency Panel Model. *American Economic Journal: Macroeconomics*, *9*, 225–253.

[3] La Grandville, O. (2017). *Economic Growth: A Unified Approach*. Cambridge: Cambridge University Press.

[4] Jones, C. I. (2003). *Growth, Capital Shares, and a New Perspective on Production Functions*. University of California Berkeley and National Bureau of Economic Research.

[5] Klump, R., McAdam, P. & Willman, A. (2007). Factor Substitution and Factor Augmenting Technical Progress in the US. *Review of Economics and Statistics*, *89*, 183–192.

Performance Comparison of Industry Clusters: Canonical Correlation Analysis vs Data Envelopment Analysis

Eva Štichhauerová[1], Miroslav Žižka[2]

Abstract. The paper evaluates the performance of companies in two industries - automotive and textile in 2019 and 2020 concerning their membership in a cluster organisation. The surveyed firms in both sectors were divided into three groups. The first group included member companies of the cluster organisation. Their performance is expected to be higher than the other two groups due to their direct involvement in cluster activities. The second group includes companies operating in the cluster's region. In this case, it can be assumed that these firms could benefit from the positive externalities of the existing cluster organisation. Their performance could thus be better than that of the third group of companies operating in other regions already too far from the cluster. Two alternative methods, Data Envelopment Window Analysis and Canonical Correlation Analysis, were used to assess the performance of the companies. Both tools can handle more inputs and outputs, but they work differently. This paper compares the results obtained using both methods and discusses their advantages and disadvantages.

Keywords: Canonical correlation analysis, Data envelopment analysis, Window analysis, technical efficiency, cluster organisation.

JEL Classification: C32, C61, L25, L62, L67
AMS Classification: 62H20, 90B90, 90C90

1 Introduction

Clusters are considered one of the key instruments of innovation and industrial policy in the Czech Republic. In particular, clusters, as entities linking businesses, research organisations, universities and government institutions, should promote efficiency, productivity and innovation. As a result, they strengthen the competitiveness of enterprises and regions [10], which is why the government strongly supports clusters under operational programmes. The cluster support through the first operational programme, Industry and Enterprise, was launched in 2004. The oldest cluster (Czech Machinery Cluster) was established in the Czech Republic in 2002. According to our research, 114 cluster organisations have been founded in the Czech Republic since 2002 (the primary data source was the public register and the keyword "cluster"). Of these, 16 clusters no longer exist, and another 47 are inactive. This means they have zero turnover, and their websites either do not exist or contain outdated information. Thus, 51 clusters can currently be described as active, regardless of their establishment year. With the establishment year of 2010 and earlier, 30 clusters are registered. This means that the oldest clusters have been operating in the Czech Republic for 20 years, and we can consider them mature in terms of their life cycle.

In the literature, many studies support but also question the expected effects of clustering on firm performance and competitiveness. In some cases, the positive impacts of clustering may appear only after a considerable time lag. For example, a study [2] showed the strongest impact of clustering in the cork industry only after more than 20 years of cluster existence. For this reason, we decided to assess the performance of two major Czech clusters founded in 2006. The first – the automotive industry – represents a growing industry that is a significant part of the Czech economy. The second – textiles – is a traditional industry that has undergone a severe decline since 1990. We intend to compare to what extent clusters have affected the performance of firms in these industries. We assess the performance of firms in clusters using two different multi-criteria methods: data envelopment analysis (DEA) and canonical correlation analysis (CCA).

The paper's main objective is to assess the performance of selected Czech industry clusters. As a sub-objective, we set out to compare the results obtained by the non-parametric DEA and parametric CCA methods and discuss the advantages and disadvantages of both approaches.

[1] Technical University of Liberec, Faculty of Economics, Department of Business Administration and Management, Studentska 2, 461 17 Liberec, Czech Republic, eva.stichhauerova@tul.cz.
[2] Technical University of Liberec, Faculty of Economics, Department of Business Administration and Management, Studentska 2, 461 17 Liberec, Czech Republic, miroslav.zizka@tul.cz.

2 Theoretical background

In decision-making, we commonly encounter situations where we assess an economic system using many inputs and outputs. In this case, we choose one of the multi-criteria evaluation methods. The most commonly used method is Data Envelopment Analysis, which can divide a set of units into efficient and inefficient ones. DEA measures the relative efficiency of a group of homogeneous production units (DMUs), usually determined by a larger number of inputs and outputs. The efficiency metric is the ratio of the weighted sum of outputs to the weighted sum of inputs. Using linear programming, we can find an empirical efficient frontier that represents best-practice observations. Units that lie on this frontier are considered efficient. Units outside it are inefficient. The original 1978 Charnes model was developed under the assumption of constant returns to scale (CRS), where the efficient frontier is a conic envelope. We distinguish the basic models into input-oriented and output-oriented. The input-oriented model assumes that the evaluated unit can influence (minimise) its inputs while maintaining a given output level. In contrast, the output-oriented model varies the output volume at a given input level. In 1984, Banker extended the model to include a variable returns to scale (VRS) condition, where the efficient frontier is a convex envelope. In the calculation, the weights of inputs and outputs are optimised to be as favourable as possible for a given unit. The highest possible efficiency score is sought for each unit under evaluation using the optimal input and output weights [4, 9]. In practice, there are several modifications of these basic DEA models, a review of which is beyond the scope of this text. In the next section, we present the mathematical model of the VRS model used in our research. Here we focus on the differences from another method, canonical correlation analysis (CCA).

The author of the CCA is considered to be Hotelling, who was to publish it in 1935 [7]. CCA measures the dependence between two groups of variables, where each group consists of at least two variables. Thus, similar to the DEA method, a larger number of inputs and outputs can be used from which CCA produces composite or canonical variables. Linear combinations of several inputs or outputs form these composite input and output variables. However, there are several linear combinations. A correlation score is sought that maximises the relationship between the composite input and the composite output. A linear combination with maximum eigenvalue determined by the CCA method [4] defines the weights. That is, there are two coefficient vectors: for inputs $\boldsymbol{U} = (a_1, a_2, \ldots, a_r)$, and for outputs $\boldsymbol{V} = (b_1, b_2, \ldots, b_s)$ [4]. Each variable has a weight (a and b, respectively). The composite variable is given by the sum of the weighted scores in each set of variables; see relations (7) and (8). These weights are called functional coefficients, and their equivalent in regression analysis are "beta weights" or "pattern coefficients" in factor analysis. Composite variables derived using these "best possible" weights are called "variate scores". The squared correlation coefficient explains the proportion of variance shared by two composite variables derived from two sets of variables [7].

DEA and CCA can be used "independently" as alternative multi-criteria assessment methods or be suitably combined. CCA can help to select appropriate inputs and outputs for subsequent DEA analysis. CCA can be applied to examine the dependence or independence of the variables used as inputs and outputs and the interrelationship between inputs and outputs. CCA can also be involved in the subsequent analysis of DEA results to rate and rank the efficient units. The set of canonical weights computed in the CCA framework [4, 6] serves well for this purpose. Based on CCA, the efficiency measure can also be determined as the ratio of weighted outputs (composite output) to weighted inputs (composite input).

3 Data and methodology

The research was carried out, given the limitations of the length of the paper, in two deliberately selected industry clusters (automotive and textile), which are at the stage of maturity. This means that they were established before 2010. However, we did not monitor the performance of enterprises only in these clusters. In each industry, non-cluster enterprises served as control groups. We further divided these non-clustered enterprises into two groups. The first group consists of enterprises that operate in the same region where the cluster is located. As these enterprises operate in the same environment, we can assume that they could partly benefit from the positive externalities of the cluster. The second group is represented by enterprises operating in the same industry but in regions other than those in which the cluster is based. It cannot be assumed that the effects of the cluster would be more significant in these regions. Thus, we monitored three groups of enterprises in each industry: cluster members (CLU), non-cluster members in the same region (REG) and non-cluster members in other regions (CZE). Performance was then compared between these groups. Given the above cluster effects, we made an assumption (1) about the performance of the firms in each group.

$$CLU > REG > CZE \tag{1}$$

Business performance was assessed on a multi-criteria basis using two inputs and two outputs. Equity and foreign capital were chosen as inputs, while outputs were sales of own products and services and economic value added (EVA). We calculated EVA using the methodology of the Ministry of Industry and Trade. The source of the accounting data was the commercial database MagnusWeb [3]. The period under review includes the years 2019 and 2020. For 2021, there are not yet sufficient accounting statements available, as companies usually have a deadline of 12 months after the end of the accounting period to publish their financial statements according to the Accounting Act. In the group of cluster organisations (denoted CLU), we monitored the performance of 10 automotive and 11 textile firms. There were 17 automotive and 51 textile companies in the second group (REG) operating in the same region as the cluster. The third group (CZE) contained 63 automotive and 178 textile firms from other regions.

In the first phase, we established pure technical efficiency scores for all companies within both industries. Due to the use of panel data, we applied a radial input-oriented DEA Window Analysis model, assuming variable returns to scale (VRS) and a window length of one year. The calculations were performed in the OSDEA-GUI software. For each group of firms (CLU, REG, CZE), we determined the average efficiency score of the two adjacent windows in which the firm was represented.

Specifically, we assume *N* units - firms ($n = 1, ..., N$) that use *r* inputs to produce *s* products in time period *T*, where $t = 1, 2,...$ T. That is, a given sample contains *N* x *T* observations. Relations (2) and (3) characterise the vectors of inputs and outputs for a particular DMU_n^t.

$$\boldsymbol{x}_t^n = (x_{1t}^n, x_{2t}^n, \dots, x_{rt}^n)' \tag{2}$$

$$\boldsymbol{y}_t^n = (y_{1t}^n, y_{2t}^n, \dots, x_{st}^n)' \tag{3}$$

Window analysis starts at time k ($1 \le k \le T$) with width w ($1 \le w \le T - k$) and has *N* x *w* observations. The input matrix X_{kw} and the matrix of outputs Y_{kw} are given by relations (4) and (5); see [8].

$$X_{kw} = \left(x_k^1, x_k^2, \dots, x_k^N, x_{k+1}^1, x_{k+1}^2, \dots, x_{k+1}^N, \dots, x_{k+w}^1, x_{k+w}^2, \dots, x_{k+w}^N\right) \tag{4}$$

$$Y_{kw} = \left(y_k^1, y_k^2, \dots, y_k^N, y_{k+1}^1, y_{k+1}^2, \dots, y_{k+1}^N, \dots, y_{k+w}^1, y_{k+w}^2, \dots, y_{k+w}^N\right) \tag{5}$$

By fitting these input-output matrices to the VRS input-oriented model (6), see, e.g. [1], we obtain the results of the DEA window analysis. For each unit, we get $w(T - w + 1)$ efficiency results (i.e., two values in our case). We compute the two values' arithmetic mean to interpret the results better.

$$\begin{gathered} \min_{\theta,\lambda} \theta = \theta_{k_{wt}}^{'} \\ \theta x_t^{'} - X_{kw}\lambda \ge 0 \\ Y_{kw}\lambda - y_t^{'} \ge 0 \\ \lambda_n \ge 0 \ (n = 1,2, \dots, N\ x\ w) \\ \sum_{n=1}^{N} \lambda_n = 1 \end{gathered} \tag{6}$$

A canonical correlation analysis was applied in the second stage for the same inputs (X_1, X_2) and outputs (Y_1, Y_2). We performed the canonical analysis for both observation periods. For both groups of composite variables, the coefficients *a* and *b* are sought such that the canonical variables *U* (7) and *V* (8) for all units exhibit the maximum pairwise correlation coefficient (9) [5]. There are several possible linear combinations between composite variables and inputs or outputs. We selected the canonical variables with the strongest correlation coefficient. In the next stage, the canonical scores for all units (enterprises) had to be determined.

$$U_{1t} = a_1 Y_{1t} + a_2 Y_{2t} \tag{7}$$

$$V_{1t} = b_1 X_{1t} + b_2 X_{2t} \tag{8}$$

$$KOR(U,V) = \frac{KOV(U,V)}{\sqrt{VAR(U)VAR(V)}} \tag{9}$$

Where: *KOR(U, V)* … coefficient of correlation between the composite input and the composite output; *KOV(U, V)* … covariance *U* and *V*, *VAR(U)* … variance *U*; *VAR(V)* … variance *V*.

We used the weights obtained by CCA to calculate the individual values of composite inputs *Z* (10) and composite outputs *W* (11). Standardisation of the variables was performed using standard deviations. The efficiency measure *R* is defined as the ratio of *W* and *Z* (12) [4].

$$W_{nt} = \frac{a_1 Y_{1t}}{S_{Y1t}} + \frac{a_2 Y_{2t}}{S_{Y2t}} \tag{10}$$

$$Z_{nt} = \frac{b_1 X_{1t}}{S_{X1t}} + \frac{b_2 X_{2t}}{S_{X2t}} \tag{11}$$

$$R_{nt} = \frac{W_{nt}}{Z_{nt}} \tag{12}$$

Subsequently, we examined the difference in DEA efficiency scores using ANOVA analysis, the respective Games-Howell test (depending on the Levene test for homogeneity of variances). Finally, we compared the ranking of the units according to the efficiency measured by the DEA and the CCA methods. The strength of the ranking dependence was measured using the Spearman coefficient and the corresponding significance level.

4 Research Results

Table 1 shows the pure technical efficiency scores according to the VRS DEA model for both industries in the years 2019-2020. At first glance, it is clear that the considered relation (1) certainly does not hold in the automotive industry. We performed ANOVA analysis to check whether the differences are significant in the textile industry (and for completeness also for the automotive industry). Respectively, in 2019, the non-parametric Games-Howell test was used because of the heterogeneity of variances, see Table 2. The results of the analyses show no significant difference in efficiency scores between groups of firms concerning cluster membership or cluster region.

Industry	**Automotive**		**Textile**	
Group	**2019**	**2020**	**2019**	**2020**
CLU	0.4313	0.4563	0.5633	0.5966
REG	0.4855	0.4281	0.5245	0.4567
CZE	0.4968	0.5481	0.5060	0.4178

Table 1 Average pure technical efficiency scores by enterprise group by year (VRS DEA Model)

Industry, Year	**Levene's test statistics (significance)**	**ANOVA F value, G-H test (significance)**
Automotive 2019	0.692 (0.503)	0.250 (0.780)
Automotive 2020	0.494 (0.612)	1.707 (0.187)
Textile 2019	3.395 (0.035)	G-H, CLU – REG: 0.886 G-H, CLU – CZE: 0.721 G-H, REG – CZE: 0.930
Textile 2020	0.449 (0.639)	2.106 (0.124)

Table 2 Results of Levene's test, ANOVA and Games-Howell tests for efficiency scores according to the DEA model

In the next phase, composite inputs and outputs for both industries were created using CCA. The weights of the inputs and outputs are shown in Table 3. Subsequently, we calculated the efficiency scores *R* and sought to determine whether its level was related to the membership of firms in the cluster or in the region. The significance of differences was tested using ANOVA, as Levene's test did not show a violation of the homogeneity of variance assumption (see Table 4).

Industry, Year	Z_{nt}		W_{nt}	
	V_{1t}	V_{2t}	U_{1t}	U_{2t}
Automotive 2019	-0.141	-0.919	-0.967	0.275
Automotive 2020	-0.239	-0.873	-0.926	0.195
Textile 2019	-0.546	-0.670	-0.885	0.251
Textile 2020	-0.602	-0.608	-1.009	0.264

Table 3 Weights of composite inputs *Z* and composite outputs *W* determined by CCA

Based on the results of the ANOVA analysis, we can conclude that the results of the efficiency score *R* are not statistically significantly influenced by the membership of the enterprise in the cluster organisation or in the region where the cluster operates. In this respect, the results are identical to the DEA analysis.

Industry, Year	Levene's test statistics (significance)	ANOVA F value (significance)
Automotive 2019	0.955 (0.385)	0.896 (0.412)
Automotive 2020	1.291 (0.280)	0.468 (0.628)
Textile 2019	1.395 (0.250)	0.061 (0.941)
Textile 2020	0.875 (0.418)	0.005 (0.995)

Table 4 Results of Levene's test and ANOVA for efficiency scores according to the CCA model

In the last step, we compared the ranking of firms obtained by DEA and CCA analyses using Spearman's correlation coefficient (see Table 5). Based on Table 5, we found a significant ranking correlation for three of the four sets examined. The only exception was in 2020 in the case of the textile manufacturing industry. We can conclude that DEA and CCA give, to some extent, identical results regarding the ranking of firms according to their efficiency scores.

Industry, Year	Coefficient
Automotive 2019	0.467**
Automotive 2020	0.587**
Textile 2019	0.138*
Textile 2020	-0.012 (sig. 0.850)

Table 5 Spearman's Rank Correlation Coefficient (** = sig. at 0.01; * = sig. at 0.05)

5 Conclusions

Based on the analyses conducted, it can be concluded that there was no significant effect of the membership of enterprises in a cluster organisation on their efficiency scores. There was also no evidence that the positive externalities resulting from the existence of a cluster organisation in a given region would increase the efficiency of non-member firms compared to their competitors from more distant regions. These findings are valid for both the automotive and textile industries studied and in both periods. In the case of the textile industry, although the absolute differences between groups of enterprises are more pronounced, the differences are not statistically significant. In this industry sector, however, it is possible that over a longer time horizon, the differences in technical efficiency become more pronounced and significant. Developments in 2020 suggest this to some extent.

As regards the comparison of the two methods used to measure the efficiency of companies - DEA and CCA – we can conclude that both can be complementary to a certain extent. For example, CCA can be used to triangulate to increase the validity of the research. Non-parametric DEA appears to be a more straightforward method in terms of computational complexity and interpretation of results. In the case of CCA, there are usually several linear combinations for each input and output, from which the best one has to be selected for further analysis. CCA is a

parametric method that uses functions to estimate a stochastic efficient frontier. In contrast, DEA models the efficient frontier empirically. Using CCA, weights of the individual inputs and outputs can be obtained similarly to DEA, but these weights can take on negative values. This situation then results in negative composite inputs and negative composite outputs. The computed efficiency score in the case of CCA can thus take on a negative value, too, which causes some interpretation difficulties.

Therefore, the applications of CCA described in the literature [4, 6] tend to use canonical correlation for post-optimisation analysis of results obtained by DEA. The reason is that the CCA method allows better smoothing of the effective frontier for selected DMUs, especially for the efficient ones. Therefore, it can be considered an alternative to DEA models for measuring super-efficiency.

Acknowledgements

Supported by the project of the Internal Grant Competition of the Faculty of Economics, the Technical University of Liberec, "The Impact of Industry Clusters on the Performance and Sustainability of Business Activities".

References

[1] Asmild, M., Paradi, J. C., Aggarwall, V. & Schaffnit, C. (2004). Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry. *Journal of Productivity Analysis*, 21(1), 67–89.

[2] Branco, A. & Lopes, J.C. (2018). Cluster and business performance: Historical evidence from the Portuguese cork industry. *Investigaciones de Historia Económica*, 14(1), 43–53. https://doi.org/10.1016/j.ihe.2016.05.002.

[3] Dun & Bradstreet Czech Republic. (2023). *Magnusweb: Komplexní informace o firmách v ČR a SR*. [Online]. Available at: https://magnusweb.bisnode.cz [cited 2023-06-27].

[4] Friedman, L. & Sinuany-Stern, Z. (1997). Scaling units via the canonical correlation analysis in the DEA context. *European Journal of Operational Research*, 100(3), 629–637. https://doi.org/10.1016/S0377-2217(97)84108-2.

[5] Ozturk, E. & Bal, H. (2017). Ranking the Airports with Data Envelopment Analysis and Canonical Correlation Analysis. *Gazi University Journal of Science*, 30(2), 237–245.

[6] Park, K. S., Lee, K. W., Park, M. S. & Kim, D. (2009). Joint use of DEA and constrained canonical correlation analysis for efficiency valuations involving categorical variables. *Journal of the Operational Research Society*, 60, 1775–1785. https://doi.org/10.1057/jors.2008.136.

[7] Thompson, B. (1984). *Canonical Correlation Analysis: Uses and Interpretation.* London: SAGE Publications.

[8] Yang, H.-H. & Chang, C.-Y. (2009). Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms. *Telecommunications Policy*, 33(1–2), 98–108. https://doi.org/10.1016/j.telpol.2008.11.001

[9] Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking* (Vol. 213). Springer International Publishing. https://doi.org/10.1007/978-3-319-06647-9.

[10] Zizka, M. & Rydvalova, P., eds. (2021). *Innovation and Performance Drivers of Business Clusters: An Empirical Study.* Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-79907-6.

Czech Republic in the Euro Area: A Two-Country DSGE Model

Patrik Šváb[1]

Abstract. This paper examines the interaction of the Czech economy (CR) and its Euro Area (EA) counterpart and the potential impacts of the monetary union entry on the Czech economy. For that purpose, I derive and calibrate a two-country DSGE model, as a merge of two basic open-economy New Keynesian models with price rigidities, supposing that the CR as a small economy is dependent on the EA, but not vice-versa. The model provides a basic tool for an impulse-response analysis with various types of shocks, originating in both economies, including the EA-CR spillovers. In the second model setup, with a potential adoption of the common currency, the results of the analysis show that without an independent monetary policy, domestic inflation would become significantly more volatile, whereas the variation in the output gap does not change. Considering a conventional intertemporal loss function, the adoption of the euro would be welfare-decreasing according to the model.

Keywords: Czech Republic, Euro Area, Two-Country DSGE

JEL Classification: E12
AMS Classification: 91B51

1 Introduction

There has been extensive research on currency unions since the topic of optimum currency areas was pioneered by Mundell [6]. Subsequently, economists enriched the theory of currency areas by criteria that a union has to fulfill to ensure that a single currency adoption will be profitable for their members. A series of applied studies followed and investigated if the adoption of the euro brings gains or losses for the economy. This paper aims to contribute to this literature by developing a basic DSGE framework that enables to map the most important linkages between the Czech and the Euro Area economies.

To my knowledge, there are not many papers that investigate welfare problems linked to a single currency adoption using (two-country) DSGE models: [1] investigate this topic for Spain, [3] for Poland, and [5] and [4] for the United Kingdom. Andrés et al. [1] conclude that inflation differentials (Spain-EA) would have been lower without the euro. Gradzewicz and Makarski [3] summarize that the euro adoption in Poland would increase output volatility, decrease inflation volatility and generate welfare losses overall. Moons [5] projects a 44% welfare loss for the UK and a 13% benefit for the EA from a potential euro accession. Lama and Rabanal [4] estimate that the euro adoption in the UK would increase welfare by 0.9%; however, it would bring a -2.9% net loss during periods with financial turbulences. It is worth mentioning that Slanicay [7] tackles the topic indirectly by analysing the asymmetry of shocks between the Czech Republic and the Euro Area and develops a two-country DSGE model for these blocks.

This paper could contribute to the literature by replicating the above-mentioned results for the Czech Republic. It is organized as follows. Section 2 briefly outlines features of the two-country DSGE model. Section 3 calibrates parameters used in the model. Section 4 shows how shocks are propagated through the blocks and spilled over to the Czech economy. Section 5 models the potential common currency adoption and section 6 concludes.

2 Model Derivation

This section shows how all the Czech and Euro Area economic agents interact in the two-country DSGE model framework. First, I use a basic New Keynesian small open economy model with price stickiness à la Calvo inspired by Smets and Wouters [8] and further adapted by Van Tran [9]. Despite the simplified forms of some functions, this model is used in certain versions by central banks to capture basic macroeconomic relationships. Therefore, I consider it suitable for the analysis that follows. Next, I extend this one-country setup to a two-country version. Parameters used in the equations are explained in Section 3.

[1] Prague University of Economics and Business, Faculty of International Relations, Department of International Economic Relations, nám. W. Churchilla 4, 130 67 Prague 3, svap02@vse.cz

Each economy is characterized by the interaction of households, firms, and central banks with specific optimization goals and stabilization objectives.

A representative **household** maximizes its utility

$$\sum_{t=0}^{\infty} \beta^t U(C_t, L_t), \tag{1}$$

following the utility function:

$$U(C_t, 1 - L_t) = \frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\phi}}{1+\phi}, \tag{2}$$

where C_t represents consumption and L_t hours worked. They encounter a budget constraint in the following form:

$$P_t C_t + Q_t B_t = B_{t-1} + W_t L_t + T_t, \tag{3}$$

where B_t represents the amount of bonds and their price Q_t is dependent on the interest rate: $Q_t = \frac{1}{1+i_t}$. W_t stands for the hourly wage and T_t are the transfer payments.

A **firm** in the domestic economy produces a differentiated good j following the production function:

$$Y_t(j) = A_t L_t(j), \tag{4}$$

where A_t represents technological progress.

The monopolistic firm decides the price of their product respecting the pricing mechanism à la Calvo. $1 - \theta$ refers to the probability that the firm changes the price of the product from $P_{D,t-1}$ to $P^*_{D,t}$.

It faces the following optimization problem:

$$\max_{P^*_{D,t}} \mathbb{E} \sum_{k=0}^{\infty} \theta^k [Q_{t,t+k}(P^*_{D,t} Y_{t+k|t} - \Psi_{t+k}(Y_{t+k|t}))], \tag{5}$$

with respect to the (domestic and foreign) demand for the firm's production at price $P^*_{D,t}$:

$$Y_{t+k|t} = \left(\frac{P^*_{D,t}}{P_{D,t+k}}\right)^{-\epsilon} (C_{t+k} + C^*_{t+k}) \tag{6}$$

for $k \in \mathbb{N}_0$, where $Q_{t,t+k}$ is the stochastic discount factor, $\Psi(.)$ is the cost function, and $Y_{t+k|t}$ stands for the production at time $t + k$, considering that firm changed the price to $P^*_{D,t}$ at time t.

Opening up the economy to the world and introducing the **international trade** brings new features to the model. Terms of trade are defined as the ratio of price levels as follows:

$$S_{i,t} = \frac{P_{i,t}}{P_{D,t}}. \tag{7}$$

Assuming the purchasing power parity condition:

$$P_{i,t} = P^i_{i,t} ER_{i,t}, \tag{8}$$

$ER_{i,t}$ is the nominal exchange rate of the currency of country i with respect to the domestic currency.

In the **equilibrium** on the market with goods produced by the firm j, the production must satisfy both the domestic and foreign demand (country i):

$$Y_t(j) = C_{D,t}(j) + \int_0^1 C^i_{Z,t}(j) di. \tag{9}$$

After a series of derivations, we can obtain the **Phillips curve**:

$$\pi_{D,t} = \beta \mathbb{E}\, \pi_{D,t+1} + \kappa_\alpha \tilde{y}_t \tag{10}$$

and the dynamic **IS curve**:

$$y_t = \mathbb{E}\, y_{t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}\, \pi_{t+1} - \rho) - \frac{\alpha\omega}{\sigma}\, \mathbb{E}\, \Delta s_{t+1}, \tag{11}$$

which are important characteristics of a small open economy. The discount factor β is defined as $\beta = \frac{1}{1+\rho}$, where ρ stands for the discount rate. π_t represents the inflation rate of all products prices and $\pi_{D,t}$ is limited to domestic products prices.

Finally, the **central bank** aims to stabilize the total price level:

$$ir_t = \rho + \Psi_\pi \pi_t. \tag{12}$$

I extend this one-country small open economy model to a two-country version for the Czech Republic and the Euro Area taking into account the following assumptions:

- Since most of the trade of the Czech Republic is realized within the EU, and the Euro Area, "the world economy" for the Czech Republic can be simplified to the Euro Area economy.
- The foreign output $\tilde{y}_t^*$ is not modeled exogenously, but it is explained in the model.
- I assume that the inflation rate π_t, output $\tilde{y}_t$, the natural level of output $\tilde{y}_t^n$, and the natural interest rate $\tilde{r}_t^n$ of the Czech Republic are influenced by those in the Euro Area, but not vice-versa.

Then, the system of equations in log-deviations from the steady state will have the following form, EA variables and parameters being denoted by an asterisk (*):

$$\begin{aligned}
\tilde{y}_t &= \mathbb{E}\, \tilde{y}_{t+1} - \frac{1}{\sigma_\alpha}(ir_t - \mathbb{E}\, \pi_{D,t+1} - \tilde{r}_t^n) \\
\pi_{D,t} &= \beta \mathbb{E}\, \pi_{D,t+1} + \kappa_\alpha \tilde{y}_t \\
\tilde{y}_t &= y_t - \tilde{y}_t^n \\
\tilde{y}_t^n &= \Gamma \tilde{a}_t + \alpha \Psi \tilde{y}_t^* \\
\tilde{y}_t &= \tilde{y}_t^* + \frac{1}{\sigma_\alpha}\tilde{s}_t \\
\tilde{r}_t^n &= -\sigma_\alpha \Gamma(\mathbb{E}\, \tilde{a}_{t+1} - \tilde{a}_t) + \alpha\sigma_\alpha(\Theta + \Psi)(\mathbb{E}\, \tilde{y}_{t+1}^* - \tilde{y}_t^*) \\
\pi_t &= \pi_{D,t} + \alpha(\tilde{s}_t - \tilde{s}_{t-1}) \\
\tilde{s}_t &= \tilde{s}_{t-1} + e\tilde{r}_t - e\tilde{r}_{t-1} + \pi_t^* - \pi_{D,t} \\
\tilde{y}_t &= \tilde{a}_t + \tilde{l}_t \\
\tilde{y}_t &= \tilde{c}_t + \frac{\alpha\omega}{\sigma}\tilde{s}_t \\
\tilde{w}_t &= \sigma \tilde{c}_t + \phi \tilde{l}_t \\
\tilde{a}_t &= \rho_a \tilde{a}_{t-1} + e_a \\
\pi_t &= \tilde{p}_t - \tilde{p}_{t-1} \\
\pi_{D,t} &= \tilde{p}_{D,t} - \tilde{p}_{D,t-1} \\
ir_t &= \Psi_\pi \pi_t
\end{aligned}$$

$$\begin{aligned}
\tilde{y}_t^* &= \mathbb{E}\, \tilde{y}_{t+1}^* - \frac{1}{\sigma_\alpha^*}(ir_t^* - \mathbb{E}\, \pi_{t+1}^* - \tilde{r}_t^{n*}) \\
\pi_t^* &= \beta^* \mathbb{E}\, \pi_{t+1}^* + \kappa_\alpha^* \tilde{y}_t^* \\
\tilde{y}_t^* &= y_t^* - \tilde{y}_t^{n*} \\
\tilde{y}_t^{n*} &= \Gamma^* \tilde{a}_t^* \\
\tilde{r}_t^{n*} &= -\sigma_\alpha^* \Gamma^*(\mathbb{E}\, \tilde{a}_{t+1}^* - \tilde{a}_t^*) \\
\tilde{y}_t^* &= \tilde{a}_t^* + \tilde{l}_t^* \\
\tilde{y}_t^* &= \tilde{c}_t^* \\
\tilde{w}_t^* &= \sigma^* \tilde{c}_t^* + \phi^* \tilde{l}_t^* \\
\tilde{a}_t^* &= \rho_a^* \tilde{a}_{t-1}^* + e_a^* \\
\pi_t^* &= \tilde{p}_t^* - \tilde{p}_{t-1}^* \\
ir_t^* &= \Psi_\pi^* \pi_t^*
\end{aligned}$$

3 Calibration

Each block is characterized by specific preferences of their agents and technological aspects, which are reflected in the parameters of the model. They are summarized in Table 1.

Parameter CZ	Value	Parameter EA	Value	Meaning
σ	0.79	σ^*	0.78	Relative risk aversion coefficient
η	0.47	η^*	0.42	Elasticity of subst. between dom. and for. products
γ	4(*)	γ^*	4(*)	Elasticity of substitution between foreign products
ϕ	1.75	ϕ^*	2.4	Fritsch labour supply elasticity
ϵ	8	ϵ^*	8	Elasticity of subst. between domestic products
θ	0.63	θ^*	0.67	Price stickiness degree
β	0.9975	β^*	0.9975	Discount factor
α	0.35	α^*	0.4(*)	Economic openness
Ψ_π	2(**)	Ψ^*_π	1.38	Taylor monetary policy rule
ρ_a	0.7(**)	ρ^*_a	0.54	Autoregressive parameter (technological shock)
ρ_c	0.95(*)	ρ^*_c	0.95(*)	Autoregressive parameter (cost-push shock)
ρ_d	0.58	ρ^*_d	0.70	Autoregressive parameter (demand shock)

Table 1: Parameters for the Czech Republic and the Euro Area

Note that if not indicated otherwise, the parameters were calibrated according to [7]. Parameters marked by (*) were adapted from [9] and parameters marked by (**) were used according to [2].

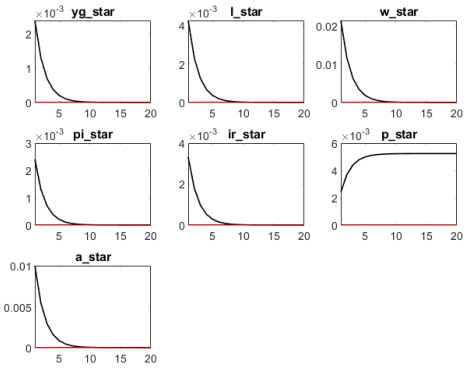
Rest of the parameters can be calculated as follows: $\omega = \sigma\gamma + (1-\alpha)(\sigma\eta - 1)$, $\sigma_\alpha = \frac{\sigma}{1+\alpha(\omega-1)}$, $\lambda = \frac{(1-\beta\theta)(1-\theta)}{\theta}$, $\kappa_\alpha = \lambda(\sigma_\alpha + \phi)$, $\Theta = (\sigma\gamma - 1) + (1-\alpha)(\sigma\eta - 1)$, $\Gamma = \frac{1+\phi}{\sigma_\alpha+\phi}$, and $\Psi = \frac{-\Theta\sigma_\alpha}{\sigma_\alpha+\phi}$ (analogically for EA parameters).

4 Shock Analysis

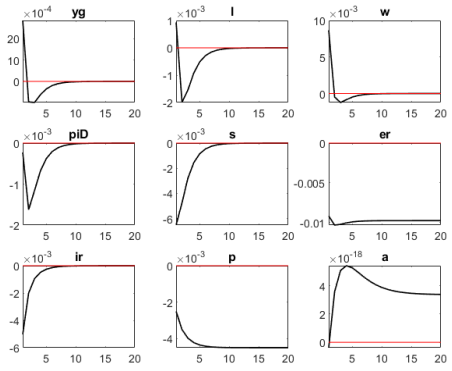
This section focuses on shocks originating in the Euro Area (EA), their propagation within this block, and the spillovers to the Czech economy (CR). All of the presented figures below depict the magnitude of change of variables (on the Y-axis) after a shock equal to an increase by 0.01 (1%). The response is simulated for 20 periods (on the X-axis). Abbreviations of the variables respect their symbols in the equation summary (see Section 2).

4.1 Technological Shock in the Euro Area

A positive technological shock in the EA should temporarily increase the natural output and, thus, narrow the output gap, see Figure 1a. As a consequence, hours worked and the hourly wage temporarily increase. EA technological improvement is also spilled over to the CR economy through increased output (see Figure 1b).



(a) EA Variables



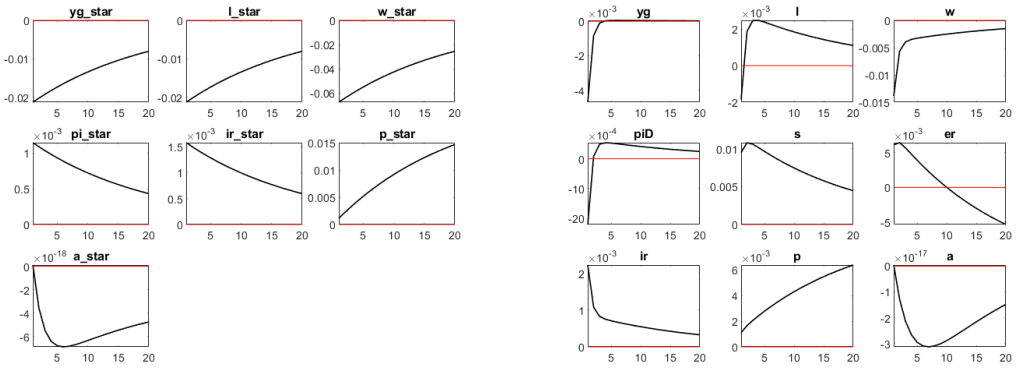
(b) CR Variables

Figure 1: Impulse-Response Functions to EA Technological Shock

4.2 Cost-Push Shock in the Euro Area

A cost-push shock in the EA causes, by definition, a raise in the price level, which is depicted for the EA economy in Figure 2a. It is accompanied by a drop in output. There is also a raise in the price level and a drop in output in

the CR economy, but with a smaller magnitude (see Figure 2b).

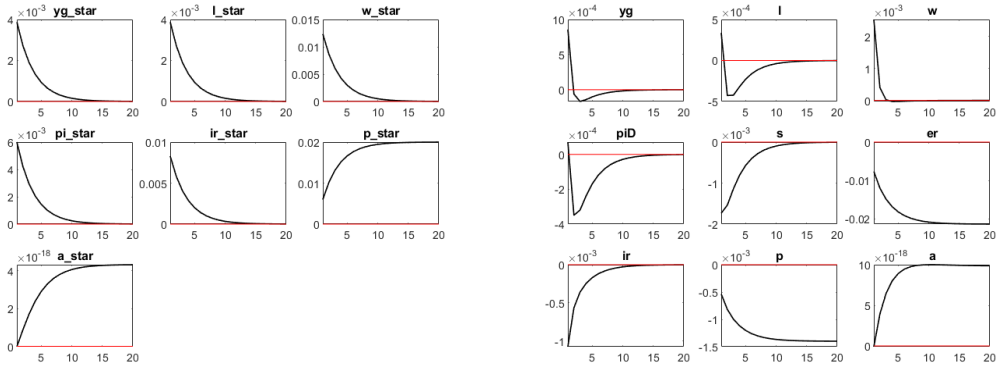


(a) EA Variables (b) CR Variables

Figure 2: Impulse-Response Functions to EA Cost-Push Shock

4.3 Demand Shock in the Euro Area

A positive demand shock in the EA block increases the output in both blocks, but with a lower magnitude in the CR block (see Figure 3a and Figure 3b). The same holds for hours worked and wages. The EA price level and the inflation rate increase. The exchange rate of the CR appreciates. The EA interest rate needs to be increased in reaction to the increased EA inflation rate.



(a) EA Variables (b) CR Variables

Figure 3: Impulse-Response Functions to EA Demand Shock

5 Common Monetary Policy

In another version of the model, I analyse the potential adoption of the common currency in the Czech Republic. In this setup, I work with two additional assumptions:

- The Czech Republic adopts the monetary policy of the Euro Area. Therefore, the autonomous monetary policy rule disappears from the model.
- With a common currency, these two blocks have no exchange rate, which is also eliminated from the model.

As a result, observed variables have different volatility under the two regimes. The variances of the key macroeconomic variables are summarized in Table 2:

We can observe that the output gap marks similar volatility under both regimes. Wage is slightly more volatile under the common monetary policy, whereas labour and terms of trade are slightly less volatile. However, the inflation rate is more volatile under the common monetary policy. Following the conventional intertemporal loss function as a weighted sum of the variances in inflation and output $L_t = \text{var}[\pi_t] + \beta_y \text{var}[y_t]$ (as in [5]), welfare losses are bigger under the common monetary policy, irrespective of the output-weight stabilization coefficient β_y.

Variable	Meaning	Independent MP	Common MP
$\tilde{y}_t$	Output gap	0.0042	0.0042
π_t	Inflation rate	0.0010	0.0057
$\tilde{s}_t$	Terms of trade	0.0025	0.0023
$\tilde{l}_t$	Labour	0.0042	0.0041
$\tilde{w}_t$	Wage	0.0182	0.0185

Table 2: Difference in Volatilites under Independent vs. Common Monetary Policy

6 Conclusion

Welfare gains and losses from a potential euro accession have been scarcely analysed through the lens of two-country DSGE models composed from a specific country and the Euro Area block. There has not been any direct contribution for the Czech Republic. Therefore, this paper aims to fill this knowledge gap. The two-country DSGE model developed in this paper is merged from two basic New-Keynesian open economy models. In the two-country setup, the foreign output is not modelled exogenously, but explained within the model. I also assumed that the key macroeconomic Euro Area variables influence the Czech economy but not vice versa. The world economy for the Czech Republic is simplified to the Euro Area. After the model is calibrated, impulse-response functions are calculated. It is shown that technological, cost-push, and demand shocks originated in the Euro Area are spilled over to the Czech economy, but with a lower magnitude.

In the alternative model setup, the volatility of variables is observed after a potential common monetary policy adoption. It is shown that the variance of the majority of main macroeconomic variables does not change significantly, nor does the output gap. The inflation rate would, however, become significantly more volatile. Considering the conventional loss function, the euro adoption would result in welfare decreasing. This result might be altered using different model specifications. It is also important to highlight that more economic and political criteria are crucial to consider while deciding about the monetary union entry. This might be part of future research. Regarding the potential future extension of this framework, the model could be extended to a third block, introducing the world economy apart from the Czech Republic and Euro Area.

Acknowledgements

The paper was supported by institutional support of the Faculty of International Relations, Prague University of Economics and Business.

References

[1] Andrés, J., Hurtado, S., Ortega, E. & Thomas, C. (2010). Spain in the euro: A general equilibrium analysis. *Journal of the Spanish Economic Association*, 1(2), 67–95.

[2] Brázdik, F., Hlédik, T., Humplová, Z., Martonosi, I., Musil, K., Šestořád, T., Tonner, J., Tvrz, S. & Žáček, J. (2020). The g3+ Model: An Upgrade of the Czech National Bank's Core Forecasting Framework. *CNB Working Paper Series*, 7, 1–58.

[3] Gradzecwicz, M. & Makarski, K. (2013). The Business Cycle Implications of the Euro Adoption in Poland. *Applied Economics*, 45(17), 2443–2455.

[4] Lama, R. & Rabanal, P. (2014). Deciding to enter a monetary union: The role of trade and financial linkages. *European Economic Review*, 72, 138–165.

[5] Moons, C. (2013). Losses from Membership in EMU: an Estimated Two-Country DSGE Model. *Applied Economics Quarterly*, 59(1), 27–61.

[6] Mundell, R. A. (1961). A Theory of Optimum Currency Areas. *The American Economic Review*, 51, 509–517.

[7] Slanicay, M. (2011). Structural Differences and Asymmetric Shocks between the Czech Economy and the Euro Area 12. *Review of Economic Perspectives*, 11(3), 168–191.

[8] Smets, F. & Wouters, R. (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, 1(5), 1123–1175.

[9] Van Tran, Q. (2019). *Makroekonomické modely pro měnovou analýzu*. Prague: Oeconomica.

Underwriting and Investment Efficiency in the Czech Life Insurance Sector: A Two-stage DEA Window Analysis Approach

Petra Tisová[1], Martin Flegl[2]

Abstract. Regarding the long insurance contracts´ duration, the investment activities mean the most important source of the revenues of the life insurers. However, the insurers must invest with respect to the Prudent Person Principle, which means investing only in instruments whose risks are clear, respect the duration of the insurance policies and are in the best interest of beneficiaries and/or policyholders. Therefore, having a confidence in the insurer(s) seems essential, especially due to the act of the delaying of the insurance payment in the case of the covered loss from the premium settlement. Considering the relatively small Czech business market, the insurers oriented on the life insurance are supposed to prove their investment effectiveness to operate on the market in the future. So, a two-stage Window Analysis Data Envelopment Analysis model was constructed to investigate the underwriting and investment efficiency of 23 Insurers. For this purpose, data from the Czech Insurance Association were used covering a period from 2012 to 2021.

Keywords: Czech Republic, Data Envelopment Analysis, Investment efficiency, Life insurance, Underwriting efficiency.

JEL Classification: C44, C61, G22
AMS Classification: 90-08, 90C05, 91B32

1 Introduction

The gross written premium obtained from the life insurance in the Czech Republic is approximately twice lesser than from the non-life insurance [9]. This fact is contrary to the situation in the developed countries of the European Union, where the considerable importance of the life insurance is perceived in general [10]. Nowadays, the insurance companies operating in the life insurance segment in the Czech Republic push hard through this state, because the potential of the market is tempting [9].

While the non-life insurance sector deals with bigger incertitude and can be perceived as volatile, as it is necessary to cope with the losses which would have the potential to be so called longed-tail, the life-sector is narrower. In general, the life insurance represents the stabilizing element for the commercial insurers, because the possibility of the catastrophic realization is by its nature significantly limited. Due to this fact, the statistical models are formally based on the two-piece normal distribution [19]. The life insurers briskly surpassed its primary intention to propose the insurance covers concerning the life issues while making the profit by this business alone and stepped into the investment markets. Nowadays, with regard to the long insurance contracts´ duration, the investment activity means the most important source of the revenues of the life insurers [2].

As the premiums of the clients constitute wherewithal for the investments of the life insurers, prudent requirements regarding the investment principles are absolutely necessary. The most concrete regulation is connected to the Solvency II, bespoke legal frame for the insurers operating in the European Union markets. Aside other realms, under the Solvency II the insurers are supposed to invest with respect to the Prudent Person Principle, which means to invest only in instruments, whose risks are clear, not deviating from the whole portfolio, which respect the duration of the insurance policies and are in the best interest of beneficiaries and/or policyholders respectively [3].

The insurance coverage works like an intangible service. Therefore, having a confidence in the insurer(s) seems to be essential, especially due to the act of the delaying of the insurance payment in the case of the covered loss from the premium settlement. While the financial funds for the investments are lower by both the lesser attractiveness of the life insurance in comparison to the non-life insurance and the relatively small Czech business market,

[1] Prague University of Economics and Business, Faculty of Finance and Accounting, Department of Banking and Insurance, Winston Churchill square 4, 130 67 Prague 3, Czech Republic, petra.tisova@vse.cz, ORCID: 0000-0003-3227-9598.
[2] Tecnologico de Monterrey, School of Engineering and Sciences, Calle Puente 222, Coapa, Arboledas del Sur, Tlalpan, 14380, Mexico City, Mexico, martin.flegl@tec.mx. ORCID: 0000-0002-9944-8475.

the insurers oriented on the life insurance are supposed to prove their investment effectiveness to secure their future operations on the market. This situation creates an environment, where the insurers' investment efficiency is supposed to be higher compared to the underwriting efficiency. To validate such assumption, an analysis is needed to observe whether the Czech insurers manage their operations in such scenario.

For this purpose, Data Envelopment Analysis is a non-parametric benchmarking approach that has been widely used for efficiency and performance evaluations in many sectors [12], [13], [16]. In the insurance sector, for example, the DEA was applied by Eling and Schaper [11] to evaluate 970 life insurers in Europe during a period of 2002-2013; Omrani et al. [18] used a two-stage DEA model to evaluate 22 Iranian insurers; Tone et al. [20] applied a DEA model to performance evaluation of 30 insurers in Malaysia from 2008 to 2016; and Wanke and Barros [21] used a two-stage DEA approach to evaluate the efficiency of 97 regulated insurers in Brazil over 1995-2013 period. In the Czech Republic, Grmanová and Pukala [14] compared the efficiency of 17 Czech commercial insurers with 26 Polish commercial insurers in 2014 applying and a DEA model and TOBIT regression. However, the quantitative research of the Czech insurance market is very limited and there is a lack of relevant studies about the efficiency in the Czech life or non-life insurance market.

So, the objective of this article is to evaluate the underwriting and investment efficiency of 23 Czech life insurers using a two-stage Window Analysis Data Envelopment Analysis model. We work with the following research questions: RQ1: What is the average underwriting efficiency? RQ2: What is the average investment efficiency? RQ3: Do the Czech life insurers operate under the assumption of greater investment efficiency over the underwriting efficiency?

2 Materials and Methods

2.1 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric approach for evaluating a set of homogeneous decision-making units (DMU) with multiple inputs and multiple outputs [7]. Each DMU consumes m different inputs to produce s different outputs. The development of the DEA model theory started with the pioneering work of Charnes et al. [4] that assumed the constant returns to scale (CCR model). If variable returns to scale are considered, the BCC model is used [1].

In many cases, the single-stage process is not be suitable to describe more complex production processes with several sub-processes. In this case, some products can be outputs of a sub-process on the one hand, and the inputs of another sub-process on the other hand. As shown in Figure 1, we assume that each DMU_j $(j = 1,2, \dots, D)$ has m inputs x_{ij}^A $(i = 1,2, \dots, m)$ used to the stage of underwriting operations, which generates D intermediate outputs z_{dj} $(d = 1,2,.., D)$ and s outputs y_{rj}^A $(r = 1,2, \dots, s)$. Then these intermediates outputs become the inputs to the stage of investment operations with m additional inputs x_{ij}^B $(i = 1,2, \dots, m)$. Finally, Stage 2 generates s outputs y_{rj}^A $(r = 1,2, \dots, s)$ [5].

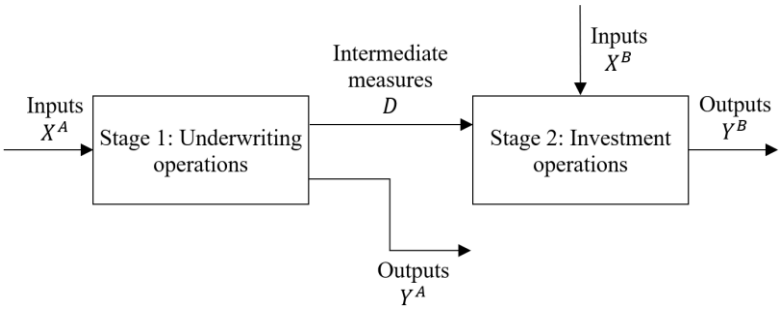


Figure 1 Two-stage model structure of the evaluation process

To measure the efficiency over a longer period, the Windows Analysis (WA) approach based on the principle of moving averages can be used [6]. The efficiency of a DMU in a particular period is compared to its efficiency in other periods, in addition to the efficiency of other DMUs. Therefore, there is $n.k$ DMU in each window, where n is the number of DMUs in each period and k is the width of each window. This feature increases the discriminatory capacity of the DEA model, as the total number of T periods is divided into series of overlapped periods (windows), each with a width $k(k < T)$ leading to $n.k$ DMUs. The first window has $n.k$ DMUs for periods $\{1, \dots, k\}$, the second period has $n.k$ DMUs and periods $\{2, \dots, k+1\}$, and so on, until the last window has $n.k$

DMUs and periods $\{T-k+1, \dots, T\}$. In total, there are $T-k+1$ separate analyses where each analysis examines $n.k$ DMUs [7].

The analysis considers two-year window ($k=2$), as only insurers that operated minimum of 2 years during the evaluated period were included in the database.

2.2 Data and Model Structure

The analysis is based on the data from the Czech Insurance Association (ČAP) from 2012 until 2021. ČAP unites the insurers generating 98% of the gross written premiums in the Czech Republic [8]. As only 2% of the insurers are not united in ČAP, we consider the Czech Insurance Association´s data representative to reflect the whole Czech insurance market. Further, we examined each Czech Insurance Association´s member, whether the Technical Account of the life insurance existed during the evaluated period. In the end, the analysis includes 23 insurers (Table 1) that operated at least two years in the market between 2012 and 2021. The two-year long period was selected to obtain complete image of the market evolution, as the Czech life insurance market is characterized by many mergers and acquisitions, as well as by many new entries.

The evaluation process of the Czech life insurers is generalized into a two-stage network structure, which is composed of underwriting operations and investment operations. For the variable selection, we considered the common DEA model structures for the life and non-life insurance companies' evaluations [11], [18], [20], [21], for which we took the indicators from the combined ratio, which is the widely used indicator for assessing the profitability of the insurers [15]. So, the underwriting stage includes Premium earned (PE), Cost of claim settlements (CCS), Claims reserves change (CRC), Other technical reserves changes (OTRC) and Net operating costs (NOC) as the inputs. Beside the premium earned as a real income of the insurer, we added Premium for reinsurance (PR), which could be perceived as a pointer of the reinsurance utilization size, and the other technical reserve changes (reflecting the uncertainty). In fact, the summation of the PE, PR and the change in reserve status equal the gross written premium, the general measurement of the insurance market performance and importance [17].

For the output part of the stage, Technical account of the life insurance (TA) and % of Premium/gross ratio (PGR) were selected. Due to the fact, that the life and the non-life activities of the insurer must be kept in the separate books, we feel free to assert that the results of TA stand for the profit or the loss in each life segment. The technical account results serve as an intermediate measure entering to the investment operations stage, as these are spent to earn finance for the insurers. In addition, Stage 2 includes Costs for investment placement realizations (CIPR) as the second input. Investment incomes (II) and Other technical incomes from the life (OTI) are the generated outputs from the Investment stage of the model.

For the calculation, the MaxDEA 7 Ultra software was used, and the BCC model was applied as we consider competition among the evaluated Insurers.

3 Results

The average underwriting efficiency of the whole sector during the evaluate period is 0.875 with standard deviation (SD) of 0.137. As Figure 2 shows, the average efficiency can be considered as stable, varying between 0.839 and 0.921. Out of the 230 observations, the insurers reached the underwriting efficiency of 1.000 in 54 occasions, representing 23.48% of the sample. The best evaluated Insurer was Basler and HVP with average efficiencies of 1.000, and ČP with an average efficiency of 0.981. However, only HVP operated during the whole evaluated period. Considering this, the 2nd highest underwriting efficiency for an insurer operating the whole period reached Cardif (0.965) and the 3rd was reached by Ergo (0.961). The lowest underwriting efficiency obtained Koop (0.655) and KP (0.717), both the Insurers with highest volatility across the evaluated period (0.146 and 0.144 respectively). Table 1 summarizes the underwriting efficiencies of all evaluated life insurers.

Regarding the investment efficiency, the average efficiency of the whole sector was 0.687 with SD of 0.316. In this case, as it can be seen in Figure 3, much higher volatility can be observed compared to the underwriting efficiency. This can be linked to the financial market development. The Eurozone economy suffered a major sovereign debt crisis in 2010 resulting in the stress for the consequent years. Since the financial market is globally connected, we assume, that this shock would have affected the investment efficiency as well. This fact led to a lower efficiency compared to the underwriting stage, varying between 0.571 and 0.756. Similarly to Stage 1, out of the 230 observations, the insurers reached the investment efficiency of 1.000 in 53 occasions, representing 23.04% of the sample. The best evaluated insurer was KP (0.998) with also one of the smallest SD (0.006), while operating the whole evaluated period. The 2nd highest investment efficiency reached KOOP (0.972, with SD 0.053), followed by ČP (0.953, with SD 0.115). The lowest investment efficiency could be observed in the case of

Simplea (0.020), Basler (0.233) and HVP (0.277). In the case of Simplea, this is a new insurer operating only the last two years, which would explain low investment incomes. Alike, Basler is not well-established insurer as the company finished its operations in 2015. Table 2 presents the results of the insurers' investment efficiencies.

INSURERS	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
AEGON	0.871	0.916	0.873	0.810	0.930	1.000	-	-	-	-
ALLIANZ	0.750	0.742	0.728	0.734	0.902	1.000	1.000	0.853	0.810	0.993
AXA ŽP	0.777	1.000	0.931	0.953	0.871	0.698	1.000	1.000	-	-
BASLER	1.000	1.000	1.000	1.000	-	-	-	-	-	-
CARDIF	1.000	0.844	0.974	1.000	1.000	1.000	0.965	0.949	0.973	0.944
ČESKÁ POJIŠŤOVNA	0.990	0.930	1.000	1.000	0.964	1.000	-	-	-	-
ČPP	0.883	0.878	0.813	0.670	0.589	0.600	0.665	0.707	0.726	0.721
ČSOB POJIŠŤOVNA	0.727	0.936	1.000	0.893	0.958	0.940	1.000	0.991	0.934	1.000
ERGO	0.964	0.972	0.975	0.928	0.930	0.928	0.986	0.962	0.965	1.000
GP	0.914	0.760	0.745	0.890	1.000	0.598	1.000	1.000	0.863	1.000
HVP	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ING	0.931	1.000	-	-	-	-	-	-	-	-
KOOP	0.684	0.719	0.543	0.641	0.473	0.595	0.736	1.000	0.574	0.588
KP	0.720	0.686	0.561	0.578	0.591	0.629	1.000	0.763	0.731	0.908
MAXIMA	1.000	0.918	0.956	1.000	0.929	0.937	0.780	0.762	0.806	1.000
METLIFE	1.000	0.899	1.000	0.746	1.000	0.663	0.850	0.768	0.846	0.767
NN	-	-	0.956	1.000	0.879	1.000	0.946	0.829	0.906	0.790
POJIŠŤOVNA ČS	0.765	0.667	0.623	0.580	0.820	0.683	1.000	-	-	-
PP	-	-	-	-	-	-	0.966	0.983	-	-
SIMPLEA	-	-	-	-	-	-	-	-	0.824	1.000
UNIQA	0.889	1.000	0.865	1.000	0.635	1.000	0.839	0.851	0.845	1.000
WUST ŽP	0.961	0.961	1.000	1.000	-	-	-	-	-	-
YOUPLUS	-	-	-	-	-	-	-	-	0.846	0.830
Average all	0.885	0.886	0.871	0.864	0.851	0.839	0.921	0.894	0.843	0.903

Table 1 Stage 1 underwriting efficiency 2012-2021

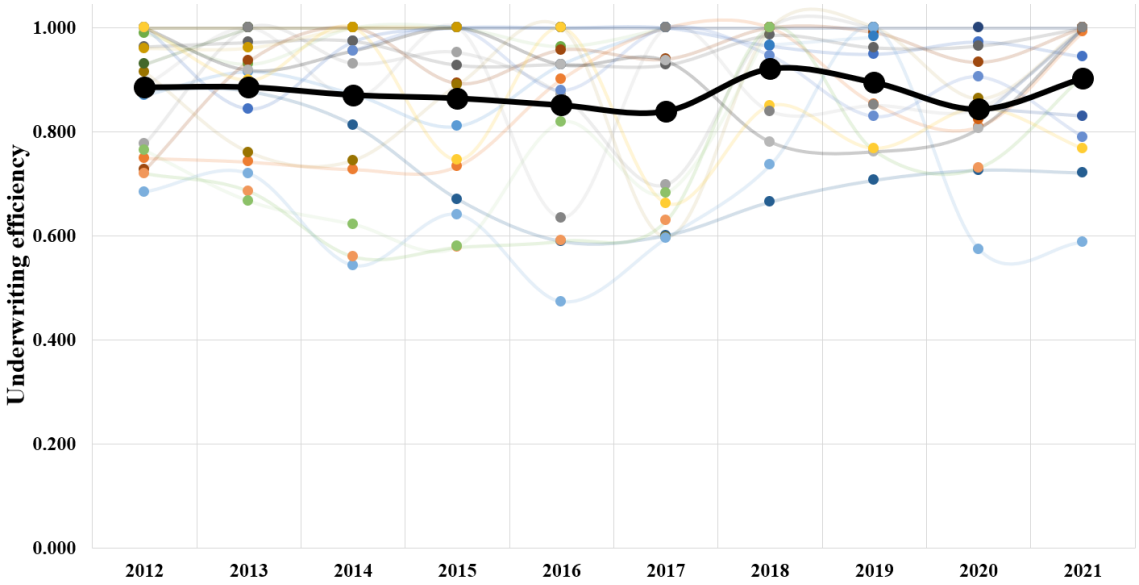


Figure 2 Stage 1 underwriting efficiency 2012-2021, dark line represents the average efficiency

INSURERS	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
AEGON	0.118	0.485	0.233	0.991	0.298	0.156	-	-	-	-
ALLIANZ	0.976	1.000	1.000	0.888	0.816	0.974	1.000	0.605	0.828	0.663
AXA ŽP	1.000	0.381	0.368	0.906	0.789	0.500	0.664	1.000	-	-
BASLER	0.089	0.239	0.532	0.071	-	-	-	-	-	-
CARDIF	0.341	0.185	0.777	1.000	0.733	0.283	0.308	0.731	0.672	1.000
ČESKÁ POJIŠŤOVNA	1.000	0.718	0.998	1.000	1.000	1.000	-	-	-	-
ČPP	0.404	0.458	0.209	0.565	0.899	1.000	0.216	0.333	0.279	0.876
ČSOB POJIŠŤOVNA	1.000	0.918	0.739	1.000	1.000	0.542	0.651	0.989	0.859	0.578
ERGO	0.364	0.226	1.000	0.537	0.722	0.198	0.929	1.000	1.000	1.000
GP	0.402	1.000	1.000	0.826	0.763	1.000	0.740	1.000	0.711	1.000
HVP	0.016	0.079	0.114	0.132	0.536	0.448	0.585	0.166	0.562	0.134
ING	0.704	0.572	-	-	-	-	-	-	-	-
KOOP	1.000	1.000	1.000	0.949	1.000	1.000	0.928	1.000	0.842	1.000
KP	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.982	1.000	1.000
MAXIMA	0.001	0.047	0.575	0.141	1.000	0.865	0.558	1.000	0.702	0.221
METLIFE	0.540	0.582	0.803	1.000	0.781	1.000	0.871	1.000	0.942	1.000
NN	-	-	0.763	1.000	0.531	0.772	0.550	0.561	1.000	0.839
POJIŠŤOVNA ČS	1.000	0.582	0.642	0.447	0.621	0.576	0.661	-	-	-
PP	-	-	-	-	-	-	0.414	0.296	-	-
SIMPLEA	-	-	-	-	-	-	-	-	0.017	0.024
UNIQA	0.578	0.449	0.670	0.243	0.356	0.079	1.000	0.355	0.880	0.672
WUST ŽP	1.000	0.932	1.000	1.000	-	-	-	-	-	-
YOUPLUS	-	-	-	-	-	-	-	-	-	-
Average all	0.607	0.571	0.706	0.721	0.756	0.670	0.692	0.735	0.735	0.715

Table 2 Stage 2 investment efficiency 2012-2021

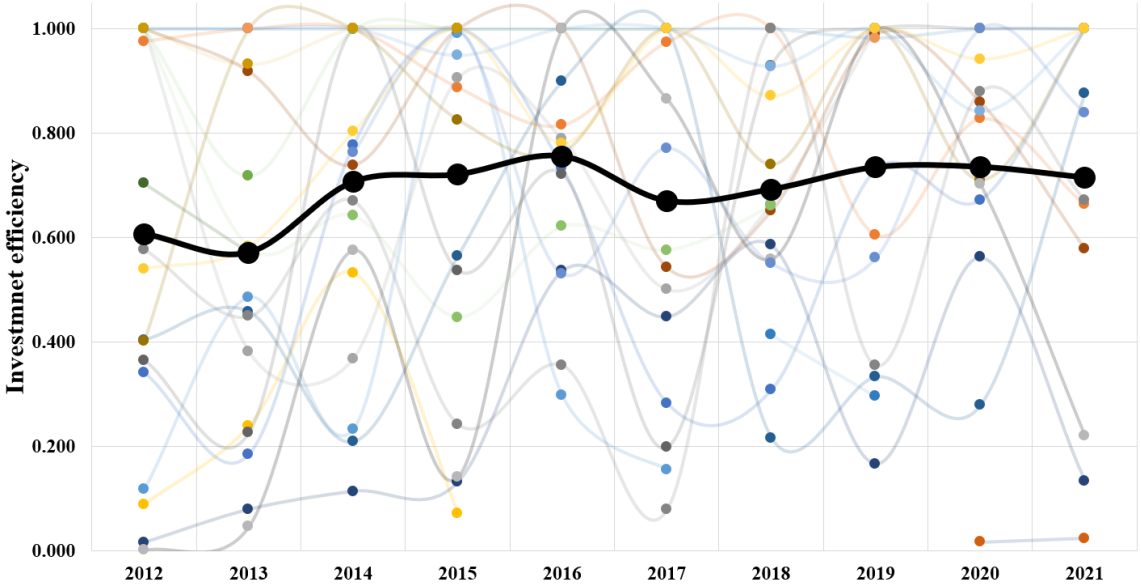


Figure 3 Stage 2 investment efficiency 2012-2021, dark line represents the average efficiency

4 Conclusion

The objective of this article was to evaluate the technical efficiency of 23 Czech life insurers during the period from 2012 to 2021. For this purpose, a two-stage Window Analysis BCC model was constructed. The technical efficiency was evaluated regarding the underwriting and investment operations using the data from the Czech Insurance Association. The analysis revealed very high underwriting efficiency with low volatility and, on the other hand, lower investment efficiency with very high volatility.

We assumed that the investment efficiency would be higher than the underwriting efficiency due to the insurers' business operations environment, which is not confirmed by the results. Afterwards we realized, that our presumption missed few important points causing this outcome. Firstly, the financial market is volatile and predisposed to the domino-effect crisis. Other co-important aspect connected to the investments is a currency rate risk. Therefore, to expect the same performance for the period of ten years is nearly impossible as no investment is not on the risk-free basis. Further, the underwriting efficiency reaches high values, because the life market is conservative and do

not offer the faulty products, and as already mentioned in the introduction, the possibility of the catastrophic realization is by its nature significantly limited. So, the obtained results can be considered valid as the calculated efficiency levels captures the nature of the life insurance market.

Also, the obtained results suggest a possible connection between both stages as, for example, KOOP and KP were evaluated within the insurers with the lowest underwriting efficiency, but among the insurers with the highest investment efficiency. Therefore, the presented analysis can be extended in this direction, i.e., to investigate relationship between the underwriting and investment efficiency.

Acknowledgements

This work was supported by the Prague University of Economics and Business, under Grant number [IGA F1/47/2022].

References

[1] 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. http://dx.doi.org/10.1287/mnsc.30.9.1078

[2] Basu, A., Bodie, Z., Drew, M. E., Kane, A. & Marcus, A. J. (2013). *Principles of Investments*. McGraw-Hill Education (Australia).

[3] Clifford Chance (2016). *Investments by Insurers under Solvency II. Briefing note*. [Online]. Available at: https://www.cliffordchance.com/content/dam/cliffordchance/briefings/2016/05/investments-by-insurers-under-solvency-ii.pdf [cited 2023-05-12]

[4] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Underwriting Research*, *2*(6), 429-444. https://dx.doi.org/10.1016/0377-2217(78)90138-8

[5] Chen, L., Lai, F., Wang, Y.-M., Huang, Y. & Wu, F.-M. (2018). A two-stage network data envelopment analysis approach for measuring and decomposing environmental efficiency. *Computers & Industrial Engineering*, *119*, 388-403. https://doi.org/10.1016/j.cie.2018.04.011

[6] Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References, and DEA-solver Software*. New York: Springer.

[7] Cooper, W., Seiford, L. M. & Zhu, J. (2011). *Handbook on Data Envelopment Analysis*, New York: Springer.

[8] ČAP (2022). *Development of the insurance market [in Czech]*. Czech Insurance Association, [Online]. Available at: https://www.ČAP.cz/statistiky-prognozy-analyzy/vyvoj-pojistneho-trhu [cited 2022-08-20].

[9] ČAP (2023). *Development of the insurance market* [in Czech]. Czech Insurance Association, [Online]. Available at: https://www.ČAP.cz/statistiky-prognozy-analyzy/vyvoj-pojistneho-trhu [cited 2023-05-12].

[10] EIOPA (2023). *Insurance statistics*. European Insurance and Occupational Pensions Authority, [Online]. Available at: https://www.eiopa.europa.eu/tools-and-data/statistics-and-risk-dashboards/insurance-statistics_en [cited 2023-05-10].

[11] Eling, M. & Schaper, P. (2017). Under pressure: how the business environment affects productivity and efficiency of European life insurers. *European Journal of Underwriting Research*, *258*(3), 1082-1094. https://doi.org/10.1016/j.ejor.2016.08.070

[12] Emrouznejad, A. & Yang, G.-L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4-8. https://doi.org/10.1016/j.seps.2017.01.008

[13] Flegl, M. & Hernández Gress, E. S. (2023). A two-stage Data Envelopment Analysis model for investigating the efficiency of the public security in Mexico. *Decision Analytics Journal*, *6*, 100181. https://doi.org/10.1016/j.dajour.2023.100181

[14] Grmanová, E. & Pukala, R. (2018). Efficiency of insurers in the Czech Republic and Poland. *Oeconomia Copernicana*, *9*(1), 71–85. https://doi.org/10.24136/oc.2018.004

[15] Leng, C.-C. (2006). Stationarity and stability of underwriting profits in property-liability insurance: Part I. *Journal of Risk Finance*, *7*(1), 38-48. https://doi.org/10.1108/15265940610637799

[16] Morán-Valencia, M., Flegl, M. & Güemes-Castorena, D. (2023). A state-level analysis of the water system management efficiency in Mexico: Two-stage DEA approach. *Water Resources and Industry*, *29*, 100200. https://doi.org/10.1016/j.wri.2022.100200

[17] OECD (2022). *Gross insurance premiums: Data*. Organization for Economic Cooperation and Development, [Online]. Available at: https://data.oecd.org/insurance/gross-insurance-premiums.htm [cited 2023-05-13].

[18] Omrani, H., Emrouznejad, A., Shamsi, M. & Fahimi, P. (2022). Evaluation of insurers considering uncertainty: A multi-objective network data envelopment analysis model with negative data and undesirable outputs. *Socio-Economic Planning Sciences*, *82*(Part B), 101306. https://doi.org/10.1016/j.seps.2022.101306

[19] Tisová, P. & Ducháčková, E. (2021). Conservative Insurance Sector: The Development under Increased Change Pressure in the Economic and Financial Environment [in Czech]. *Socio-Economic and Humanities Studies*, *13*(1), 141-152.

[20] Tone, K., Kweh, Q. L., Luc, W.-M. & Ting, I. W. K. (2019). Modeling investments in the dynamic network performance of insurers. *Omega*, *88*, 237-247. https://doi.org/10.1016/j.omega.2018.09.005

[21] Wanke, P. & Barros, C. P. (2016). Efficiency drivers in Brazilian insurance: A two-stage DEA meta frontier-data mining approach. *Economic Modelling*, *53*, 8-22. https://doi.org/10.1016/j.econmod.2015.11.005

Portfolio Cash Flow on Peer-to-Peer (P2P) Lending Platform: The Quantile Regression Approach

Petra Vašaničová[1], Marta Miškufová[2]

Abstract. Peer-to-peer (P2P) lending is a financial technology that has emerged in recent years as an alternative to traditional lending methods. Bondora is a European P2P lending platform that offers investors the opportunity to invest in consumer loans originated in Estonia, Finland, and Spain. This paper analyzes the cash flow statement of Bondora. The aim of this paper is to find out the relationship between paid principal and paid interest of Bondora platform using quantile regression on monthly data from April 2009 to April 2023. Quantile regression is a valuable statistical technique that offers advantages over traditional linear regression because it can determine whether individual percentiles of a dependent variable are more affected by independent variables than other percentiles of a dependent variable, which is then reflected in the change in regression coefficients. This study examines the performance of Bondora's loan portfolio and discusses the future of Bondora and its potential for growth and expansion. Results show that over time the influence of the paid principal on the paid interest increases. This study offers valuable insights to investors for evaluating the profitability of their investments and enables them to make informed decisions about future investments on the P2P lending platform.

Keywords: peer-to-peer lending, P2P, Bondora, cash flow, quantile regression

JEL Classification: G23, O16, C21
AMS Classification: 62G08

1 Introduction

Peer-to-peer (P2P) lending is a form of lending that enables individuals to borrow and lend money directly without the involvement of traditional financial institutions, in the P2P lending platform [36]. P2P lending platforms typically operate online and use algorithms and automated processes to match borrowers with lenders based on their creditworthiness, risk profile, and investment objectives [9]. It is important to note that P2P lending involves risks, including the possibility of default by borrowers and the lack of government-backed insurance or protection for investors.

There are many P2P lending platforms operating around the world, each with its own unique features, lending criteria, and risk profile. The well-known P2P lending platforms are Prosper (based in the United States, offers personal loand and debr consolidation loans), Zopa (based in the UK, offers personal loans, auto loans, debt consolitation loans), Mintors (European platform, allows investors to invest in loans from a variety of loan originators from around the world), Bondora (European platform, offers personal loans, business loans, and portfolio management services). It is important to conduct thorough research and due diligence before investing in any P2P lending platform because these investments can involve significant risks and potential losses. This can involve reviewing the platform's historical performance, analyzing its lending criteria and underwriting process, and evaluating the platform's risk management strategies.

In this paper, we focusses on Bondora's portfilio cashflow indicators (paid principal and paid interest) that characterize its historical performance. Portfolio cashflow refers to the net cash inflows and outflows generated by a portfolio of investments over a specific period. It takes into account all the income and expenses associated with the investments in the portfolio. The portfolio cashflow is an important metric that investors use to track the performance of their investment portfolios and to assess their ability to generate income. Investors can use portfolio cashflow analysis to make informed decisions about their investments, such as deciding which assets to buy or sell, or whether to allocate more or less capital to different investment strategies.

[1] University of Presov, Faculty of Management and Business, Department of Finance, Accounting and Mathematical Methods, 17. novembra 1, 080 01 Presov, Slovakia, petra.vasanicova@unipo.sk.
[2] University of Presov, Faculty of Management and Business, Department of Finance, Accounting and Mathematical Methods, 17. novembra 1, 080 01 Presov, Slovakia, marta.miskufova@unipo.sk.

The aim of this paper is to find out the relationship between paid principal and paid interest of Bondora platform. In the context of P2P lending, portfolio paid principal refers to the amount of principal that has been repaid by borrowers on a portfolio of loans that an investor has invested in. On the contrary, portfolio paid interest refers to the amount of interest that has been paid by borrowers on a portfolio of loans that an investor has invested in. Investors can use the portfolio paid principal and portfolio paid interest metrics to calculate their portfolio return. This can help investors assess the profitability of their investments and make informed decisions about future investments on the P2P lending platform.

2 Literature review

Cash flows have value relevance [15] and are a fundamental performance measure for a firm's valuation [25]. Operating cash flows are crucial for decision makers when they are taking decisions that are related to financing future projects and repaying debts that support business activities and enhance profitability. In addition, it helps management to take decisions that are related to dividend policy. Gomez [10] argued that operating cash flow is the most important factor in predicting the financial crisis and also helps to provide a perception of business life cycle.

Barth et al. [1] argued that the market's assessment of firms' cash flows influences the relation between price and earnings. Overall, a firm with a more reliable cash flow may show greater solvency and may be more attractive to investors [41]. The cash flow of a firm has an important meaning besides its profit. According to Nguyen and Nguyen [34], cash flow statements affect information users' decisions when corporate profit decreases.

Erich [6] determined that those who use financial statements for lending decisions should consider cash flow statements to assess the financial condition of a firm. For the lenders, sufficient and useful accounting information on the cash flow statement provided may help them to analyze and evaluate the business performance, thus, to ensure the lenders make reasonable loan decisions to avoid risks and improve the efficiency of capital lending fully and comprehensively.

Yan et al. [42] highlighted that net cash flow, as one of the most important financial and credit status of lending platforms, can play a crucial role in platform operation and gaining investors' trust. Ma and Wen [30] have also mentioned that P2P lending promoted information flow in the process of cash flow and the break of cash flow would cause the bankrupt of the platform.

Online P2P lending refers to the unsecured direct loans between lenders and borrowers through online platforms without the intermediation of any financial institutions [12], [28]. Different from traditional bank lending, P2P lending is based on individual and SMEs who are expected to raise money from the vast number of investors via the online virtual platforms. Each platform uses their own evaluation models to decide the interest rate [13]. Thus, differences in risk evaluation models make the P2P platforms have different interest intervals, which affects the P2P market volatility [7].

Several authors have investigated the factors that influence the interest rate and success rate on P2P platform. They divided these factors into internal and external. Among the internal factors they included, for example gender [17], [33], race [39], [35], credit score [14], [21], [32], [20], statement of loan [26], [18], social capital [27], [28], [16], market mechanism [31], liability [43], P2P lending platform characteristics [40], historical performance [5], appearance characteristics and living area [11]. Macroeconomic elements, such as government bond yield and unemployment level [4], [8], are considered as external influencing factors.

It follows that the existing literature on P2P lending deals with individual loans, credit and profit risk for investors [29], characteristics of borrowers. This paper contributes to the P2P lending market literature by addressing portfolio performance of specific P2P lending platform.

3 Data and Methodology

3.1 Data

We use monthly data from April 2009 to April 2023 from the Cash Flow Statement of Bondora platform database. Specifically, we use paid principal (PP) that denotes the amount of principal repaid during the period, and paid interest (PI) that denotes the amount of interest paid during the period. In the context of P2P lending, portfolio paid principal refers to the amount of principal that has been repaid by borrowers on a portfolio of loans that an

investor has invested in. Figure 1 presents the development of PP and PI in analyzed period. We see that the paid interest grows more slowly than the paid principal.

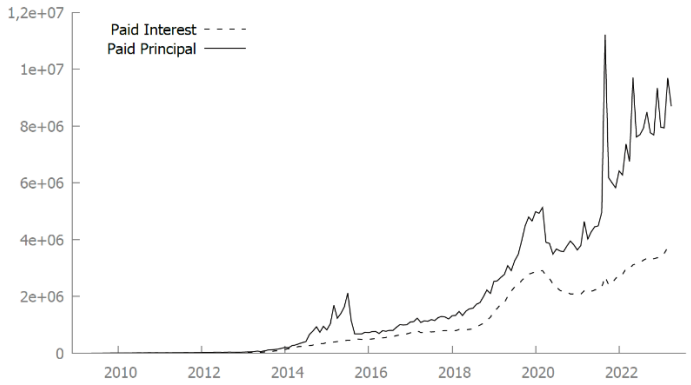


Figure 1 Development of PI and PP from April 2009 to April 2023
Source: own processing

3.2 Quantile Regression

In this Section, we describe basic concept of the quantile regression methodology, according to [19, p. 25], [38, p. 387]. In the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, ..., n, \tag{1}$$

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 $X_i = \left(X_{i1}, ..., X_{ip} \right)^T$ denotes the i-th observation and ρ_τ (defined in [22] as loss function) is considered in the form

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

with indicator function denoted by 1. Alternatively, ρ_τ may be formulated as

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

If we assume that the quantile τ of the conditional distribution of the dependent variable Y_i is a linear function of the vector of independent variables (X_i), then we can write the quantile conditional regression as [24]:

$$Y_i = \beta_0 + \boldsymbol{\beta}_\tau \mathbf{X}_{i\cdot} + \varepsilon_{i\tau}, \quad i = 1, ..., n, \tag{5}$$

A specific feature of quantile regression is that the estimated coefficients of the independent variables, $\boldsymbol{\beta}_\tau$, can be significantly different in various quantiles, which may indicate a heterogeneous conditional distribution of the dependent variable [3]. The advantage of QR is that it is the most suitable tool for modeling heteroscedastic data [19, p. 25], [22].

To meet the aim of this paper, the model for the OLS is:

$$PI_i = \beta_0 + \beta_1 PP_i + \varepsilon_i, \quad i = 1, ..., n \tag{6}$$

while for QR, we consider the model according to (5) and the sequence of estimated coefficients is from $\tau = 0.05$ to $\tau = 0.95$ by 0.05. We test the presence of heteroscedasticity by Breusch-Pagan test. If the residuals are heteroskedastic in the regression model, we use a paired bootstrap to compute *p*-values. To estimate the regression parameters of the QR model, we use the RStudio and the quantreg package, which was created according to [22], [23]. To test whether the slope coefficients of the models are identical, we use ANOVA and the anova.rq package.

4 Results and Discussion

Table 1 presents the estimates of QR and OLS models. Results of the ANOVA test detected that QR estimates significantly differ across quantiles. The regression model parameter estimates obtained using OLS were statistically significant, and the model explained up to 90.69% of the variability of the *PI*. However, we indicated the presence of heteroskedasticity, which we confirm through the Breuch-Pagan test (BP = 31.278, $p = 0.0000$). Therefore, the use of quantile regression is justified. The results of QR show that *PP* is statistically significant on each quantile level. We show that paid interest on Bondora platform is determined by paid principal. Through quantile regression, we found out which percentiles of paid interest may be more influenced by paid principal (we see high coefficients for high values of quantiles). It turns out that over time the influence of the paid principal on the paid interest increases.

Dependent variable: Paid interest (PI)				
Quantile	**Intercept**	***p*-value**	**Paid principal (PP)**	***p*-value**
0.05	-2815.35281	0.2218	0.31509	**0.0000**
0.10	-3625.41244	**0.0061**	0.36041	**0.0000**
0.15	-3583.09865	**0.0004**	0.38513	**0.0000**
0.20	-3864.38658	**0.0001**	0.40513	**0.0000**
0.25	-3765.44899	**0.0015**	0.41346	**0.0000**
0.30	-3014.27160	**0.0404**	0.41506	**0.0000**
0.35	-2499.08580	0.3228	0.42569	**0.0000**
0.40	-2304.59072	0.8151	0.43480	**0.0000**
0.45	-835.70848	0.9508	0.44259	**0.0000**
0.50	2283.98225	0.8619	0.46867	**0.0000**
0.55	28956.58938	0.0946	0.50339	**0.0000**
0.60	34160.73869	0.0622	0.52047	**0.0000**
0.65	54752.34459	**0.0022**	0.53451	**0.0000**
0.70	52459.61868	**0.0000**	0.56025	**0.0000**
0.75	75071.87719	**0.0000**	0.56450	**0.0000**
0.80	72506.10493	**0.0000**	0.58023	**0.0000**
0.85	72539.48236	0.1865	0.59492	**0.0000**
0.90	45768.31397	0.4190	0.63416	**0.0000**
0.95	27991.85148	**0.0055**	0.68509	**0.0000**
ANOVA *p*-value = 0.0000				
OLS	158600	**0.0000**	0.4187	**0.0000**
BP = 31.278 (*p*-value = 0.0000); $\boldsymbol{R^2} = 0.9069$				

Table 1 Estimates of model parameters
Source: own processing
Note: The P-values marked bold indicate the statistical significance at the significance level of 0.05.

Figure 2 presents the sequence of estimated coefficients from $\tau = 0.05$ to $\tau = 0.95$ by 0.05. Each panel represents a covariate in the model; the horizontal axes display the quantiles while the estimated effects are reported on the vertical axes [2] [38]. The horizontal black solid line parallel to the x-axis denotes zero value; the red solid line corresponds to the OLS coefficient along with the 95% confidence interval (red dashed lines). Each black dot is the slope coefficient for the quantile indicated on the x-axis with 95% confidence bands marked by grey color [37] [38]. As is stated in [2, p. 14], a joint inspection of the QR coefficients and the corresponding confidence bands, along with the OLS confidence intervals permits an understanding of whether the effect of predictors is significantly different across the conditional distribution of *PI* values compared to the OLS estimate.

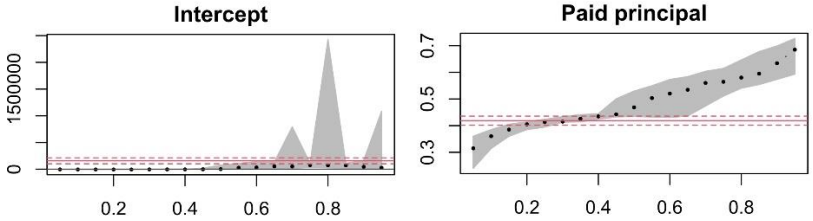


Figure 2 Estimates of model parameters by quantile level
Source: own processing

Figure 3 shows a scatterplot, OLS and QR fit for different taus. Superimposed on the plot are the $\tau = 0.05$, $\tau = 0.10$, $\tau = 0.25$, $\tau = 0.75$, $\tau = 0.90$, $\tau = 0.95$ quantile regression lines in gray, the median fit in solid blue, and the least squares estimate of the conditional mean function as the solid red line.

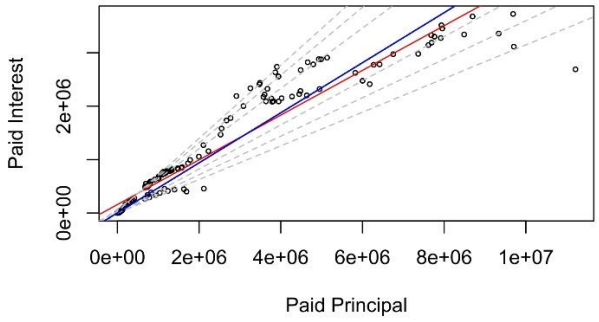


Figure 3 OLS and QR fit for different taus
Source: own processing

In practice, the following applies: As borrowers make their loan payments, a portion of each payment goes towards paying off the principal amount of the loan, while the remaining portion goes towards paying the interest charged on the outstanding loan balance. Over time, as borrowers make their scheduled payments, the amount of outstanding principal decreases, which means that the interest charged on the remaining balance also decreases. This relationship means that as more principal is paid off, the amount of interest paid on the outstanding balance decreases as well. In other words, the amount of paid interest is dependent on the amount of outstanding principal. As the outstanding principal decreases, the amount of interest charged on the remaining balance decreases, which means that the amount of paid interest also decreases. Similarly, as more principal is paid off, the amount of outstanding balance decreases, which also leads to a decrease in the amount of paid interest. Therefore, tracking both paid principal and paid interest is important for investors in order to get a comprehensive understanding of the performance of their investments on a P2P lending platform.

5 Conclusion

Bondora is well-established European P2P lending platform that was founded in Estonia in 2008. It offers a variety of loan types and features for investors. The platform operates in several European countries, including Estonia, Finland, Spain. In this paper, we examined the relationship between two portfolio cashflow indicators, i.e., paid principal and paid interest, of Bondora platform using quantile regression on monthly data from April 2009 to April 2023. Quantile regression is a valuable statistical technique that offers advantages over traditional linear regression because it can determine whether individual percentiles of a dependent variable are more affected by independent variables than other percentiles of a dependent variable, which is then reflected in the change in regression coefficients. Our results show potential for growth and expansion of Bondora platform. However, as with any investment, it is important for investors to carefully evaluate the risks and benefits of investing on the platform. In the future research, it would be interesting to use data from other platforms and conduct comparative analysis of the outcomes. Additionally, we see the opportunity to employ quantile regression to explore the relationships between various loan-related variables on the Bondora platform.

Acknowledgements

This paper was supported by the research grant VEGA no. 1/0497/21 “Risk modeling in the P2P lending market” and KEGA no. 001PU-4/2022 “Application of Modern Trends in Quantitative Methods in the Teaching of Financial and Managerial Subjects”.

References

[1] Barth, M. E., Cram, D. P. & Nelson, K. K. (2001). Accruals and the prediction of future cash flows. *Accounting Review*, *76*(1), 27–58.

[2] Costanzo, A. & Desimoni, M. (2017). Beyond the mean estimate: a quantile regression analysis of inequalities in educational outcomes using INVALSI survey data. *Large-Scale Assessments in Education*, *5*(14), 1–25.

[3] Cupák, A., Pokrivčák, J. & Rizov, M. (2016). Diverzita spotreby potravín na Slovensku. *Politická ekonomie*, *64*(5), 608–626.
[4] Dietrich, A. & Wernli, R. (2016). *What drives the interest rates in the P2P consumer lending market? Empirical evidence from Switzerland*. Social Science Electronic Publishing.
[5] Ding, J., Huang, J., Li, Y. & Meng, M. (2019). Is there an effective reputation mechanism in peer-to-peer lending? Evidence from China. *Finance Research Letters*, *30*, 208–215.
[6] Erich, A. H. (2001). *Financial Analysis Tools and Techniques: A Guide for Managers* (1st ed.). New York: Mc Graw-Hill Education.
[7] Fang, X., Wang, B., Liu, L. & Song, Y. (2018). Heterogeneous Traders, Leverage Effect and Volatility of the Chinese P2P Market. *Journal of Management Science and Engineering*, *3*(1), 39–57.
[8] Foo, J., Lim, L. H. & Wong, S. W. (2017). *Macroeconomics and fintech: uncovering latent macroeconomic effects on peer-to-peer lending*. arXiv: Econometrics.
[9] Ge, R., Feng, J., Gu, B. & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, *34*(2), 401–424.
[10] Gomez, L. (2002). Enron - A Case for Better Understanding of Cash Flows. *Business Credit*, *104*, 12–13.
[11] Gonzalez, L. & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, *2*, 44–58.
[12] Greiner, M. E. & Wang, H. (2010). Building consumer-to-consumer trust in e-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce*, *15*(2), 105–136.
[13] He, F., Li, Y., Xu, T., Yin, L., Zhang, W. & Zhang, X. (2020). A Data-Analytics Approach for Risk Evaluation in Peer-to-Peer Lending Platforms. *IEEE Intelligent Systems*, *35*(3), 85–95.
[14] Herzenstein, M., Andrews, R. I., Dholakia, U. M. & Lyandres, E. (2008). The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities. *Boston University School of Management Research Paper*, *14*(6), 1–36.
[15] Hirshleifer, D., Hou, K. & Teoh, S. H. (2009). Accruals, cash flows, and aggregate stock returns, *Journal of Financial Economics*, *91*(3), 389–406.
[16] Chen, X. R., Zhou, L. N. & Wan, D. F. (2016). Group social capital and lending outcomes in the financial credit market: An empirical study of online peer-to-peer lending. *Electronic Commerce Research and Applications*, *15*, 1–13.
[17] Chen, D. Y., Li, X. L. & Lai, F. J. (2017). Gender discrimination in online peer-to-peer credit lending: evidence from a lending platform in China. *Electronic Commerce Research*, *17*(4), 553–583.
[18] Chen, X., Huang, B. & Ye, D. (2018). The role of punctuation in P2P lending: Evidence from China. *Economic Modelling*, *68*, 634–643.
[19] Kalina, J. & Vidnerová, P. (2019). Implicitly weighted robust estimation of quantiles in linear regression. In M. Houda & R. Remeš (Eds.), *Conference Proceedings of the 37th International Conference on Mathematical Methods in Economics 2019* (pp. 25–30). České Budějovice: University of South Bohemia in České Budejovice.
[20] Kgoroeadira, R., Burke, A. & André van Stel, A. (2019). Small business online loan crowdfunding: who gets funded and what determines the rate of interest? *Small Business Economics*, *52*(1), 67–87.
[21] Klafft, M. (2009). Peer-to-peer-lending: auctioning microcredits over the internet. In A. Agarwal & R. Khurana (Eds.), *Proceedings of the International Conference on Information Systems, Technology and Management*. Dubai: IMT.
[22] Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.
[23] Koenker, R., Chernozhukov, V., He, X. & Peng, L. (2017). *Handbook of Quantile Regression*. Boca Ranton: CCR Press.
[24] Kováč, Š. (2013). Vybrané faktory predlženosti podnikov v podmienkach SR. *Forum Statisticum Slovacum*, *7*, 79–85.
[25] Larrain, B. & Yogo, M. (2008). Does firm value move too much to be justified by subsequent changes in cash flow? *Journal of Financial Economics*, *87*(1), 200–226.
[26] Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D. & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, *39*(1), 19–37.
[27] Lee, E. & Lee, B. (2012). Herding behavior in online P2P lending: an empirical investigation. *Electronic Commerce Research and Applications*, *11*(5), 495–503.
[28] Lin, M., Prabhala, N. R. & Viswanathan, S. (2013). Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science*, *59*(1), 17–35.

[29] Lyócsa, Š., Vašaničová, P., Hadji Misheva, B. & Vateha, M. D. (2022). Default or profit scoring credit systems? Evidence from European and US peer-to-peer lending markets. *Financial Innovation*, *8*(1), 1–21.
[30] Ma, B. & Wen, Z. (2016). Models, risks, and regulations of P2P lending in China. In *Proceedings of 2015 2nd International Conference on Industrial Economics System and Industrial Security Engineering* (pp. 341–348). Singapore: Springer.
[31] Ma, B. J., Zhou, Z. L. & Hu, F. Y. (2017). Pricing mechanisms in the online peer-to-peer lending market. *Electronic Commerce Research and Applications*, *26*, 119–130.
[32] Michels, J. (2012). Do Unverifiable Disclosures Matter? Evidence from Peer-to-Peer Lending. *The Accounting Review*, *87*(4), 1385–1413.
[33] Mohammadi, A. & Shafi, K. (2017). Gender differences in the contribution patterns of equity-crowdfunding investors. *Small Business Economics, 50*(2), 275–287.
[34] Nguyen, D. D. & Nguyen, V. C. (2020). The Impact of Operating Cash flow in Decision-Making of Individual Investors in Vietnam's Stock Market. *The Journal of Asian Finance, Economics and Business*, *7*(5), 19–29.
[35] Pope, D. G. & Sydnor, J. R. (2011). What's in a picture? Evidence of discrimination from Prosper.com. *The Journal of Human Resources, 46*(1), 53–92.
[36] Serrano-Cinca, C., Gutiérrez-Nieto, B. & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one, 10*(10).
[37] Vasanicova, P., Jencova, S., Gavurova, B. & Bacik, R. (2021). Cultural and Natural Resources as Determinants of Travel and Tourism Competitivenes. *Transformations in Business & Economics*, *20*(3), 300–316.
[38] Vašaničová, P. & Jenčová, S. (2022). Determinants of International Tourism Inbound Receipts: The Quantile Regression Approach. In H. Vojáčková (Ed.), *Conference Proceedings of the 40th International Conference on Mathematical Methods in Economics 2022* (pp. 386–391). Jihlava: College of Polytechnics Jihlava.
[39] Walter, T. (2008). Competition to default: Racial discrimination in the market for online peer-to-peer lending. *Business*, 1–44.
[40] Wang, Q., Xiong, X. & Zheng, Z. (2021). Platform Characteristics and Online Peer-to-Peer Lending: Evidence from China. *Finance Research Letters*, *38*, 101511.
[41] Wild, J., Subramanyam, K. R. & Hasley, R. (2004). *Financial Statement Analysis*. New York: McGraw-Hill/Irwin.
[42] Yan, Y., Lv, Z. & Hu, B. (2018). Building investor trust in the P2P lending platform with a focus on Chinese P2P lending platforms. *Electronic Commerce Research, 18*(2), 203–224.
[43] Zhou, Y. & Wei, X. (2020). Joint liability loans in online peer-to-peer lending. *Finance Research Letter, 32*, 101076.

Clustering Methods Usable in Loss Reserving in Non-Life Insurance and Their Comparison

Petr Vejmělka[1]

Abstract. Insurance companies perform loss reserving as an important part of their activities. Reserving corresponds to the estimation of an insurer's liability from future claims payments. We can find many approaches to estimating the pertinent reserves in the literature. In this article, we consider a method using state-space modeling that transforms run-off triangles, a widely considered scheme of claims, into time series with missing observations. Usually, there is more information about claims available, such as the type of claim, on the basis of which it would be possible to split these claims to several groups according to their similarity. With this approach, we would achieve greater homogeneity within the given groups, and due to this, we could also expect more accurate estimates.
The aim of this paper is to compare clustering methods applicable as part of the reserving that are available in the form of different packages implemented within the R software. This paper includes an application of unsupervised classification to claims portfolios, which were created with the use of a generator designed based on a real portfolio. Thanks to this generator, future expenses for incurred claims are also known, and it is therefore, possible to compare how accurate the reserve estimates are.

Keywords: reserve, clustering, state-space model, time-series, non-life insurance

JEL Classification: C32, C53, G22
AMS Classification: 91G05

1 Introduction

In this article, we are dealing with the estimation of reserves in non-life insurance. There are two main principles of how one can approach the task of reserving. Macro-reserving uses models that work with aggregate data in run-off triangles. The second option, which is more common in the literature recently, is micro-reserving. In contrast to macro-reserving, in the case of micro-reserving claims are modeled individually. As an alternative, a compromise that combines the simplicity of the aggregated method and the additional information known about individual claims can be considered.

The aim of this paper is to present an overview of clustering methods that can be used to divide claims in non-life insurance into several groups according to their similarities. Theoretically, this division should increase homogeneity within these clusters. Thus, the predictive abilities of methods used for reserving ought to be better in comparison to the case when all claims are considered together in just one group. In this paper, we limited ourselves only to methods that are implemented in software R in the form of packages.

There are many methods that can be used to estimate a reserve. For the purpose of this work, we decided to consider an approach transforming original claims data arranged into a form of run-off triangles into time series with missing observations. For such time series it is possible to use state-space modeling. Regarding a specific model, we consider a log-normal model presented in [5].

To have a sufficient amount of data on which to make the comparison, we decided to use the claims portfolio generator proposed by [10]. This generator was created in a way that the generated claims should reflect a real insurance claims portfolio. For a sufficiently large sample of generated portfolios, we compare how accurate the reserve estimates were, compared to the actual values for the considered clustering methods.

This paper is organized as follows. Section 2 introduces a multivariate log-normal model, which is used for reserve estimation in state-space modeling. In Section 3, we present an overview of several clustering methods and the generator. Section 4 is devoted to the comparison of results. Finally, Section 5 summarizes the conclusions achieved in this paper.

[1] MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8, vejmelp@karlin.mff.cuni.cz

2 State-Space Modeling

In this article, the estimation of future payments for already incurred claims is based on state-space modeling. For this reason, we will first introduce the linear state model.

2.1 Linear State-Space Model

In a macro-reserving approach, claims are usually aggregated to run-off triangles. One of the possibilities how to estimate the reserves is to consider state-space modeling. To be able to use such a method, firstly, it is necessary to transform the original run-off triangles into a time series form. There are three main approaches how to order the data, i.e., row-wise ordering (based on accident years), column-wise ordering (based on development years), and ordering based on calendar years. In this paper, we decided on a row-wise ordering. For data in a time-series form, it is possible to use state-space modeling. In our case, it is sufficient to limit ourselves only to a linear state-space model. This model, in its general form, is given by the following system of equations:

$$y_t = Z_t \alpha_t + \varepsilon_t, \tag{1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \tag{2}$$

where y_t is a p-dimensional observation vector at time t, α_t is a m-dimensional state vector at time t and further Z_t, T_t and R_t are matrices of parameters of types $(p \times m)$, $(m \times m)$ and $(m \times k)$. Moreover, it is assumed that $\varepsilon_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$ and $\alpha_1 \sim N(a_1, P_1)$ are indpendent, where H_t and Q_t are covariance matrices of types $(p \times p)$ and $(k \times k)$ and both a_1 and P_1 are some initial estimates. Since the data are considered in the form of time series with missing observation, Kalman smoothing is used for the estimation of unknown values, which can be further used to estimate the reserves.

2.2 Multivariate Log-Normal Model

In literature, several different models based on the state-space model have been proposed. For the purpose of the reserving and afterward comparison among considered clustering methods, a multivariate log-normal model presented in [5] will be used. This model works with the incremental claims X_{ij}, where i denotes the accident year and j denotes the development year expressing the delay with which the claims, or their parts, were paid to beneficiaries. It is assumed that these incremental claims have log-normal distribution, i.e., $X_{ij} \sim LN(\mu_{ij}, \sigma_\varepsilon^2)$. That means that after a logarithmic transformation, values $Y_{ij} = \log X_{ij}$ are assumed to be normally distributed. Hence, for values Y_{ij}, we have

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2), \tag{3}$$

where

$$\mu_{ij} = c + a_i + b_j, \tag{4}$$

and where the parameter a_i in (4) represents the row effect and b_j the column effect.

As mentioned above, it is necessary to change the double-index to a simple index that would correspond to a considered time series. In our case, we are going to assume that there is the same number of rows and colums, namely $s + 1$, labeled as $i, j = 0, \ldots, s$. The corresponding time index is calculated than as $t = i \cdot s + j$. Since we assume that our data are made up of multiple time series, we need to modify the log-normal model to its multivariate version. Let us suppose, that there are N run-off triangles with values $X_{ij}(n)$, where n denotes the nth run-off triangle for $n = 1, \ldots, N$. Then the log-normal model can be written as follows:

$$y_t(n) - y_t^0(n) = \alpha_t(n) + \varepsilon_t(n) \tag{5}$$

$$\alpha_{t+1}(n) = \alpha_{t-s+1}(n) + \eta_t(n), \tag{6}$$

where $y_t(n)$ and $y_t^0(n)$ represent the corresponding terms $Y_{ij}(n)$ and $Y_{i0}(n)$, $\alpha_t(n)$ is a state variable in the nth run-off triangle, $\varepsilon_t(n) \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon(n,n))$ and $\eta_t(n) \overset{\text{ind.}}{\sim} N(0, \sigma_\eta(n,n))$. The value $\sigma_\varepsilon(m,n)$ stands for $\sigma_\varepsilon(m,n) = \text{Cov}(\varepsilon_t(m), \varepsilon_t(n))$ and, similarly, for $\sigma_\eta(m,n)$, where $m, n = 1, \ldots, N$. Further, the subtraction of values $y_t^0(n)$ that corresponds to the subtraction of the values from the first columns of run-off triangles from the remaining values in particular rows, is done. This step is considered in order to set some initial levels in the observation equation. Therefore, these values are not estimated.

Finally, we need to adjust this model to a matrix form that would correspond to the linear state-space model defined by (1) and (2). This adjustment is crucial to be able to estimate the model using suitable software. The final form

of the model is

$$y_t - y_t^0 = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \alpha_t + \varepsilon_t, \tag{7}$$

$$\alpha_{t+1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & & & & \\ 1 & 0 & \cdots & 0 & 0 & & & & \\ \vdots & \vdots & & \vdots & \vdots & & & & \\ 0 & 0 & \cdots & 1 & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 0 & 0 & \cdots & 0 & 1 \\ & & & & & & 1 & 0 & \cdots & 0 & 0 \\ & & & & & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & & & & \\ 0 & 0 & \cdots & 0 & 0 & & & & \\ \vdots & \vdots & & \vdots & \vdots & & & & \\ 0 & 0 & \cdots & 0 & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1 & 0 & \cdots & 0 & 0 \\ & & & & & & 0 & 0 & \cdots & 0 & 0 \\ & & & & & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \eta_t, \tag{8}$$

where $y_t = (y_t(1), \ldots, y_t(N))'$, $y_t^0 = (y_t^0(1), \ldots y_t^0(1))'$, $\alpha_t = (\alpha_t(1), \ldots, \alpha_{t-s+1}(1), \ldots \alpha_t(N), \ldots, \alpha_{t-s+1}(N))'$, $\varepsilon_t = (\varepsilon_t(1), \ldots, \varepsilon_t(N))'$ and $\eta_t = (\eta_t(1), 0, \ldots, 0, \ldots, \eta_t(N), 0, \ldots, 0)'$. The covariance matrices of the residual vectors ε_t and η_t are then $\mathrm{Var}(\varepsilon_t) = H_t = \left(\sigma_\varepsilon(m,n)\right)_{m,n=1,\ldots,N}$ and $\mathrm{Var}(\eta_t) = Q_t = \left(\sigma_\eta(m,n)\right)_{m,n=1,\ldots,N}$.

Since we decided to perform a comparison of clustering methods available in R, estimation of the proposed model and reserves is being done in R as well. One of the possibilities how to perform the Kalman smoothing in R is to use the KFAS package that was created by Helske. For more information about this package, see [4]. When the model is estimated and smoothed values are calculated, these values are transformed back to the original scale. Finally, the estimates of the previously unknown values are summed up. This sum corresponds then to the desired reserve.

3 Overview of Clustering Methods

In this section, we present the clustering approaches that will be compared in the numerical study. For each method, the main idea of how the clustering is done, is mentioned. More detailed information can be found in the original articles.

3.1 Anticlust

First of the functions that can be used for clustering is a function called *balanced_clustering* from the package *anticlust*, see [8]. One of the specifics of this function is that it creates clusters that are of equal size. The principle of this method is as follows. Firstly, the centroid of all data, which is defined as the mean vector of all columns of a feature matrix, is computed. After identification of the centroid, an observation that is the most distant from the centroid is assigned to a cluster together with its $\frac{n}{k} - 1$ nearest neighbors, where n is the number of observations and n is the number of clusters, that should be created. This proximity is based on the Euclidean distance. This procedure is repeated with the remaining observations until each of the observations is clustered.

3.2 K-Means

Another option for clustering data is the use of the k-means algorithm. One of the functions that perform k-means clustering is a function *Kmeans* from the *stats* package. This function is based on an algorithm proposed in [3]. In the beginning, k observations are chosen at random, each of them representing a cluster. According to the distance between these chosen observations and the remaining ones, clusters of unassigned observations are determined. After this initial division into groups, the calculation of cluster centroids and reassignment of observations to new clusters, based on the distance from the centroids, is repeated until the clusters do not change.

3.3 Model Based Clustering

A completely different approach is considered by the function *Mclust* in the package *mclust*. This approach is called model-based clustering. The main idea of this method is a consideration that data come from a mixture of

densities, where each of the k clusters is modeled by the Gaussian distribution. For different covariance matrix parametrizations, the maximum likelihood method is used to fit these models. Based on an information criterion, the best model is chosen. After that, the most probable cluster for each observation is predicted. More information about the latest version of this package can be found in [9].

3.4 Clustering Large Applications

The last considered method is called Clustering Large Applications and is usually abbreviated as CLARA, which was proposed in [6]. Unlike other methods that may have problems with the large number of observations, this method enables us to perform clustering even on large datasets. In this case, a small sample from the data is taken, and the Partitioning Around Medoids (PAM) algorithm is applied. This algorithm generates an optimal set of medoids for the considered sample. Medoid is a term used for clusters, whose average disparity over all objects in the cluster is minimal. Firstly, a suitable initial set of medoids is searched, and in the second step exchanges between the initial medoids and the observations are provided as long as there is a decrease in the objective function. There are multiple implementations of this algorithm in R. We decided to use package *ClusterR* with the function *Clara_Medoids*.

4 Comparison

4.1 Generation

To be able to compare the mentioned methods, it is necessary to have suitable insurance data. For this purpose, we chose the generator of insurance claim portfolios, whose implementation in R is part of [10]. We decided to consider a ten-year period during which claims occurred and were subsequently settled. We have partially modified the original portfolios to make the data fit the purpose of the comparison. Each claim is then settled within ten years since the time of its occurrence. Therefore, in the case of development triangles, we consider $s = 9$. Table 1 shows the form in which the data are being considered.

Id	Type	Number of payments	Accident month	Month of payment	Payment	Total claim
1	2	1	1	42	6,175	6,175
2	1	2	1	9	3,609	6,521
2	1	2	1	16	2,912	6,521
3	2	2	1	7	64,314	122,690
3	2	2	1	12	28,376	122,690
4	5	1	1	20	2,330	2,330

Table 1 Illustration of data

Each line corresponds to one payment for an insurance claim. These claims are distinguished by their identification numbers listed in the *Id* column, are of a given type, and are settled within the given number of payments captured in the column *Number of payments*. Claims of five different types are considered in total. This is followed by the month when the damage occurred and the month when the corresponding payment was made. For simplicity, the months are numbered from 1 to 228, i.e., there are 19 years during which the claims are settled. The last two columns present the payment amounts and total claims amounts. Claims and payments are considered in CHF.

We generated a total of 500 insurance claim portfolios. The number of claims varies between individual portfolios, but on average, there are 30,000 of them in each portfolio. For each dataset, a value of how much money is paid out for claims incurred during the first ten years over the following nine years is computed. When reserving, one does not know these future payments, and it is necessary to estimate the sum that corresponds to the reserve. Thanks to the considered generation, we know the actual values to which the reserves should be close. Thus, it is possible to compare whether any of the clustering methods improves the predictive capability of the state-space model or not.

In order to make a relevant comparison, we had to decide what number of clusters to consider. We have come to the conclusion that the most favorable results for the considered data are achieved with two clusters. Therefore, we limited ourselves to this case only. The variables, on the basis of which payments were clustered, are the number of payments for the given claim, the number of months between the occurrence of the claim and the corresponding payment, the amount of the payment, the total claim and its type, which was transformed into five binary variables. Without this adjustment, it would not be possible to include this variable in clustering.

As part of the comparison, we have also included two simpler methods that divide claims based on their amount. In the first case, claims were split into two halves, below and above the median. In the second case, the dividing line was set as the 80% quantile. The first group consisted of more minor claims, the second group contained fewer major claims.

As an example of the output of the state-space model, we present Figure 1. It consists of two graphs of smoothed values for one randomly selected portfolio that was clustered using the CLARA method. The black curves correspond to the known values that were used for prediction construction, the red dashed lines represent the projections.

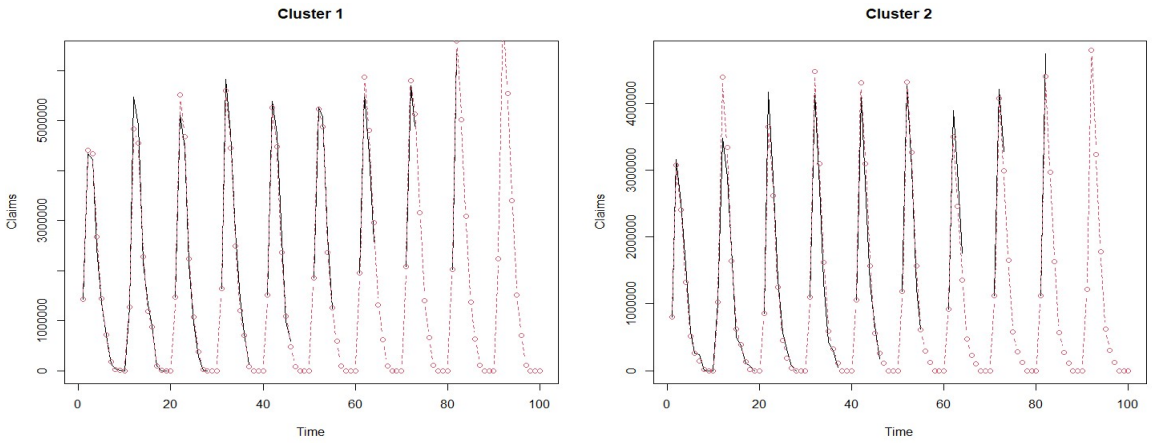


Figure 1 Illustration of the original and smoothed values for a randomly selected portfolio

4.2 Results

We can now proceed to the numerical results. As a measure of the success of the forecasts, we decided to use the sums of the absolute deviations of the reserves from the actual values. These sums are available in Table 2. The method with the best result corresponding to the lowest value in the table is highlighted. For a better idea, we supplement this table with a boxplot comparison, which can be seen in Figure 2.

Without clustering	50/50	80/20	Anticlust	Kmeans	Mclust	Clara
$2.131 \cdot 10^9$	$2.366 \cdot 10^9$	$2.754 \cdot 10^9$	$2.452 \cdot 10^9$	$3.517 \cdot 10^9$	$2.752 \cdot 10^9$	$2.074 \cdot 10^9$

Table 2 Comparison of different methods - sum of absolute deviations from actual values

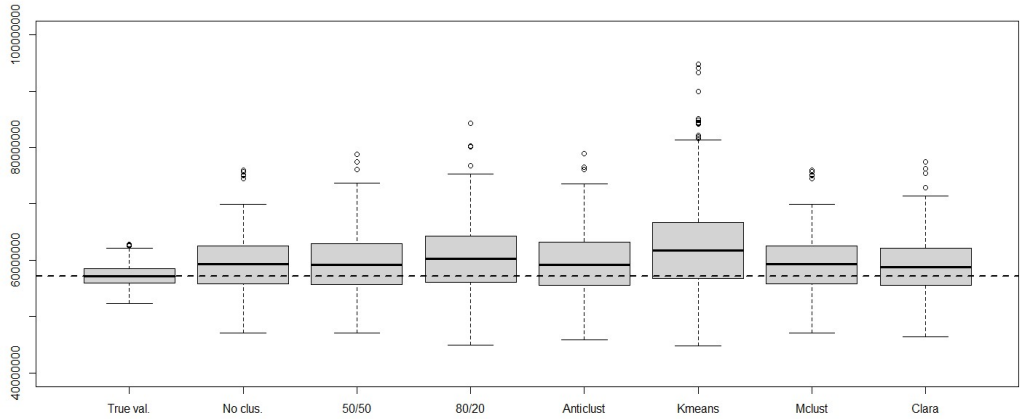


Figure 2 Boxplots of the reserves and corresponding actual values

It is clear that for the vast majority of clustering methods, the corresponding estimates are less accurate than in the case of the non-clustered variant. For some approaches, such as K-means, this deterioration is significant.

However, one of the clustering methods achieved better results, namely, CLARA. We can compare the number of simulations when reserves obtained with the use of CLARA were more accurate than reserves calculated without any clustering. In that case, the ratio would be 256:244 in favor of CLARA.

5 Conclusions

The aim of this article was to find out whether it is advantageous to consider clustering in the framework of reserving in non-life insurance. We intended to verify whether the assumption that, in the case of ensuring greater homogeneity within the individual damage groups, the predictive ability of the models will increase, is valid in practice. We performed this verification on the data obtained using the insurance claims generator. Although these claims are simulated, they should correspond to real claims. For clustering purposes, we considered four clustering methods supplemented by two straightforward approaches. The reserving was made using state-space modeling, specifically, the log-normal model was chosen. For clustering, its multidimensional form was considered.

It turns out that using clustering does not necessarily lead to better results. Only one of the methods, namely CLARA, achieved more accurate predictions. It should be noted that results can vary widely depending on what data are used. However, for portfolios that are composed of a more diverse range of claims, clustering could be of greater benefit.

Acknowledgements

This paper was supported by the grant 19-28231X provided by the Czech Science Foundation.

References

[1] Brockwell, P. J. & Davis, R. A. (1991): *Time Series: Theory and Methods* (Second Edition). Springer-Verlag, New York, ISBN 978-0-387-97429-3.

[2] Durbin, J. & Koopman, S. J. (2002): A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3), 603–615.

[3] Hartigan, J. A. & Wong, M. A. (1979): Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28, 100–108.

[4] Helske, J. (2017): KFAS: Exponential family state space models in R. *Journal of Statistical Software*, 78(10), 1–38.

[5] Hendrych, R. & Cipra, T. (2020): Applying state space models to stochastic claims reserving. *ASTIN Bulletin*, 51(1), 267–301.

[6] Kaufman, L. & Rousseeuw, P. J. (1990): *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, ISBN 0-471-87876-6.

[7] Mouselimis, L. (2023): ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering. *R package version 1.3.1*.

[8] Papenberg, M. & Klau, G. W. (2021): Using anticlustering to partition data sets into equivalent parts. *Psychological Methods*, 26(2), 161–174.

[9] Scrucca, L. et al. (2016): Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 289–317.

[10] Wang, M. & Wüthrich, M. (2022): Individual claims generator for claims reserving studies: Data Simulation.R. *SSRN Electronic Journal*.

The Influence of Influencers: Assessing the Impact of Influencer Marketing on Brand Awareness

Lukáš Veverka[1]

Abstract. This article investigates the impact of influencer marketing on webpage visits, which can serve as a proxy for brand awareness. Using a dynamic time series methodology, the study evaluates the effect of influencer programs on organic webpage visits by estimating both the long-run propensity and impact propensity. The study employs Fourier transformation to estimate seasonality throughout the year and finite distributed lag to assess the immediate impact of influencers. It also looks for patterns within the month, such as the payday effect. The empirical data used for this research was obtained from Google Analytics for a Fast Moving Consumer Goods (FMCG) company that sells clothes. The findings suggest an initial increase in webpage visits, but statistical significance cannot be established.

Keywords: Fourier transformation, Dynamic time series, Data-driven marketing

JEL Classification: C22, M31
AMS Classification: 91B84

1 Introduction

Affiliate marketing and influencer marketing are two growing industries that have recently gained much attention. In affiliate marketing, a company pays affiliates to direct people to their website with the aim of generating sales. Affiliates promote the company through their own personal networks, websites or social media platforms. Conversely, influencer marketing involves a company working with an individual with a significant online presence and influence over a particular target audience. The influencer advertises the company's products or services to their audience through sponsored posts, videos, and other forms of content, emphasising building trust and increasing brand exposure.

This article aims to investigate the impact of influencers on the webpage visits, which serves as an approximation of brand awareness. It emphasizes the importance of considering both the short-term and long-term effects of marketing campaigns to enhance their effectiveness and promote corporate growth. By evaluating the impact of influencer programs, businesses can improve their marketing efforts and increase the effectiveness of such activities.

2 Literature Review

Several studies have been conducted on using Instagram influencers as affiliate partners due to their increasing popularity. Lou and Yuan [5] explored whether influencers can enhance brand awareness and purchase intentions and identified several critical elements for success, including the influencer's dependability and attractiveness. Building on this study, Casaló et al. [3] emphasized the importance of an influencer's creativity and distinctiveness in becoming an opinion leader and positively influencing customer behavioural intentions. Sokolova and Kefi [7] conducted a study on audience engagement that is parasocial in nature and found that physical beauty does not have much impact on interactions. De Veirman et al. [4] also examined the effect of an influencer's follower count and the number of accounts they follow. They discovered that an influencer's popularity may suffer if they follow a small number of accounts.

3 Analyzed Data Sample

For this empirical study, data for a fast-moving consumer goods (FMCG) firm that sells clothing in Central and Eastern Europe (CEE) was gathered from Google Analytics. The daily pageviews are the dependent variable. The data sample consists of 1161 observations spanning October 28, 2018, and December 31, 2021. Table 1 includes a sample of the data.

[1] University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vevl00@vse.cz

Pageviews	Day year	Month course	Discount 25 %	After XMAS	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Influencer campaign
8141	301	28	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
9485	302	29	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
10189	303	30	0	0	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
10651	304	31	0	0	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
9053	305	1	0	0	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
7599	306	2	0	0	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
7437	307	3	0	0	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
8491	308	4	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
11521	309	5	0	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
12797	310	6	0	0	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Table 1 Data sample of the first ten rows emerging into the model

Additional variables were created based on the pageviews date to analyse the influencer campaign's effect on pageviews. *After XMAS*, *Month course*, *Day year*, and dummy variables for the days of the week are among the variables derived from the date. These factors offer crucial details regarding the seasonality and brand awareness tendencies that may affect the firm.

The start of the discount event, as announced by the influencers, is represented by the dummy variable *Influencer campaign*. Another dummy variable, *Discount 25 %*, depicts dates associated with an annual special discount event that often significantly influences the page views. The *After XMAS* variable denotes the period following the Christmas shopping frenzy, which typically follows the weekend before Christmas. This variable provides important data regarding the trends in Christmas shopping patterns.

4 Methodology

4.1 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, and econometrics. In this paper, we will use it for the approximation of seasonality. It works similarly to Taylor polynomial, which combines power series to approximate a difficult function around a certain point. Fourier transformation is based on combining trigonometric functions, which are able to converge to any periodical function. An extreme example is square or sawtooth waves which are, in fact, angular. Fourier transformation is able to converge even to them [6].

Fourier transformation is based on combining trigonometric series. The principal thought is based on assembling an anisochronous (i.e. having different frequencies) 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:

$$f(t) = \beta_0 + \sum_{n=1}^{\infty} A_n \sin(\omega n t) + \sum_{n=1}^{\infty} B_n \cos(\omega n t), \tag{1}$$

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 a year) and t is a linearly increasing vector containing the same number of elements as observations and its values are in the range of 0 and length of time periods (e.g. years). Finally, A_n and B_n are coefficients creating a spectrum of the function [2]. These coefficients are then estimated with OLS (ordinary least squares) so the resulting function would approximate the seasonality of a given time series.

4.2 Finite Distributed Lag

In a finite distributed lag (FDL) model, the influence of one or more variables (such as x_t on y_t) with a temporal delay is considered. There are several causes for the lagged effect of the change in x_t, including biological, economic, or behavioural ones (such as the relationship between inflation and central interest rates). An example of an FDL model of order k (allowing up to k delays) is Equation (2).

$$y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \ldots \delta_k x_{t-k} + u_t, \tag{2}$$

where δ_0 denotes the instantaneous change in y_t caused by an increase of x_t by one unit. The impact propensity is commonly used to describe it. The changes in y_t that occur in one, two, and k periods following the temporary modification are denoted by the variables δ_1, δ_2, and δ_k, respectively. The lag distribution is produced by graphing all of the δ parameters as a function of lag (k). It shows how a brief rise in x_t affects y_t dynamically. A distribution like this has no assumptions and is not constrained (unlike the generalized gamma distribution) [1]. The long-term change in y_t is represented by the sum of the estimated δ parameters ($\sum_{i=0}^{k} \delta_k$) for the current and delayed effects of change in x_t. The long-run propensity, or LRP, is frequently the most intriguing quantity examined in distributed lag models. Despite the possibility of multicollinearity making individual δ parameter estimations inaccurate, the estimation of the LRP is frequently accurate [8].

5 Results

We include Fourier transformation in our model because we assume that pageviews show a significant amount of seasonality. A set of control variables (see Table 1) are then introduced, including the number of days in a week. The influencer campaign is finally introduced with four lags. The resulting model can be shown in the format shown below:

$$y_t = \beta_0 + f(\phi_n, \theta_n, \kappa_m, \tau_m, yr_t, mo_t) + \sum_{i=2}^{n} \beta_i x_{it} + \sum_{l=0}^{4} \delta_l c_{t-l} + u_t, \tag{3}$$

where y_t are the pageviews, $f(\phi_n, \theta_n, \kappa_m, \tau_m, yr_t, mo_t)$ represents the Fourier transformation and is described in equation (4), the influencer campaign's time series is represented by the variable c_t. Lagged version is identified as c_{t-l}. The remaining variables, x_i, are primarily used to ensure that the model is complete.

$$f(\phi_n, \theta_n, \kappa_m, \tau_m, yr_t, mo_t) : \sum_{n=1}^{O_{yr}} \phi_n \sin(\omega n yr_t) + \sum_{n=1}^{O_{yr}} \theta_n \cos(\omega n yr_t) + \sum_{m=1}^{O_{mo}} \kappa_m \sin(\omega n mo_t) + \sum_{m=1}^{O_{mo}} \tau_m \cos(\omega n mo_t) \tag{4}$$

In equation (4), $\phi_n, \theta_n, \kappa_m, \tau_m$ stand for parameters to be estimated during OLS estimation, yr_t, mo_t represent the number of the day in a year and the number of the day in a month, respectively. O_{yr} and O_{mo} are the number of orders of the Fourier transformation. Finally, ω, as described in section 4.1, represents how many times the seasonal wave appears in one time period – in this case, it is 2π since the wave is only one.

The evaluation of the interactions between intra-month patterns and yearly seasonal patterns takes place at the beginning of the modelling phase. In order to do this assessment, a grid search must be conducted for various Fourier transformation order combinations. It is considered that both year and month seasonal patterns can have up to 50 orders of Fourier transformation – this gives us 2500 combinations to test. The Akaike information criterion (AIC) is then used to evaluate the results. This evaluation reveals that the first-order Fourier transformation, which has the lowest AIC, is the method that best approximates the intra-month pattern. However, an ANOVA test shows that the intra-monthly pattern is insignificant (see Table 2). This pattern is thus ignored in the study's next stages.

Model	RSS	Res.Df	Sum of Sq	F	Pr(>F)
Including month pattern	1.3629×10^{10}	1117			
Without month pattern	1.3647×10^{10}	1119	-18377325	0.7531	0.4711

Table 2 Anova output for models with and without the intra-month pattern.

The intra-month pattern's lack of significance makes it conceivable to reconsider the ideal number of orders for the Fourier transformation. The lowest Akaike information criterion (AIC), which is attained at 21 orders, is shown in Figure 1. Figure 2 shows the estimated seasonality and also shows how the seasonality might change if different orders were chosen. We specifically chose to compare it with an order of 10 since it produced the lowest AIC before another fall (see Figure 1). Most pageviews take place right before Christmas. To better understand this pattern, a dummy variable (*After XMAS*) that denotes the conclusion of the Christmas shopping season was added

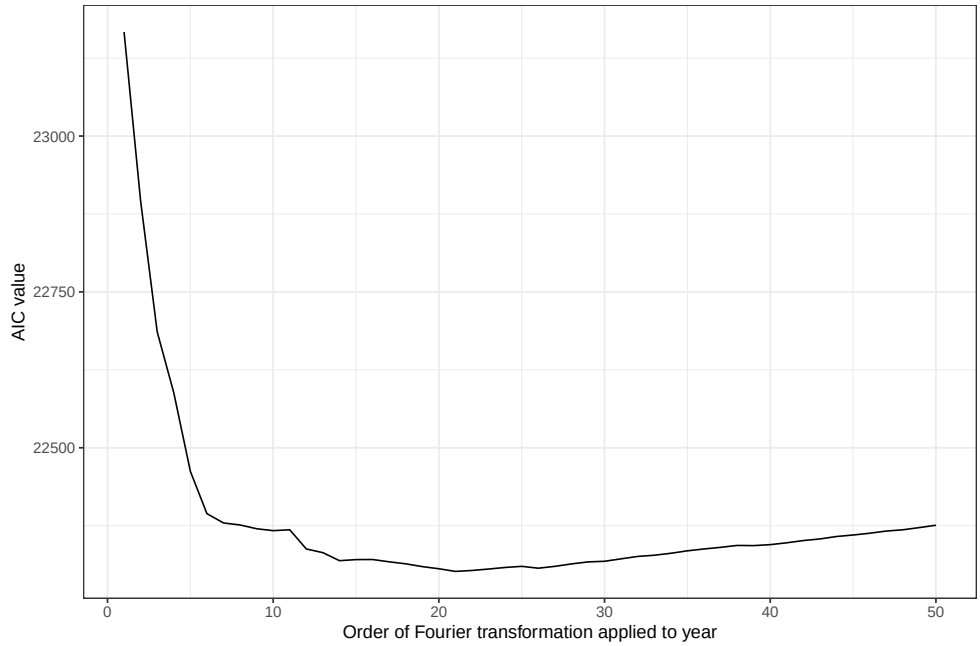


Figure 1 Dependency of the AIC value on the order of Fourier transformation applied to variable year. The minimal AIC is at order 21.

to the data. The resolution, which can vary from year to year, often occurs the week following the weekend before Christmas.

According to the Breusch-Godfrey test for correlation of order up to 7 (potential weekly serial correlation), the model does contain considerable first-order autocorrelation. As a result, the lagged y_t term is necessary to include. Autocorrelation is then no longer a problem for such a model. The model also has a reasonable fit of $R^2_{adj} = 0.8465$.

The model's most notable results are the estimated parameters for the influencer campaign's impact – δ parameters in equation (3). The initial response to the campaign is an increase of 2480.3 pageviews which is the estimation of the δ_0 known as the impact propensity. This implies that the discount promotion offered by influencers directly impacts brand awareness the day it is introduced. The following δ parameters show that the campaign also significantly affects the following day. The campaign appears to have a negative impact on the third and fourth days. This can be explained as a saturation of the marketing effect.

The Long Run Propensity (LRP), which has been determined to be positive (LRP = 6223.7), determines the campaign's total impact. This indicates that even while pageviews may start to fall toward the end of the campaign, it still helps the company overall because it generates an additional 6223.7 pageviews. The expected evolution of pageviews following the start of influencer marketing is shown in Figure 3.

However, the uplift in pageviews caused by the influencer campaign is statistically inconclusive because the effect is insignificant (based on the significance testing in Table 3).

Model	RSS	Res.Df	Sum of Sq	F	Pr(>F)
Including influencer campaign	5.8388×10^9	1046			
Without influencer campaign	5.8825×10^9	1051	-43711337	1.5661	0.1669

Table 3 Anova output for models with and without the influencer campaign.

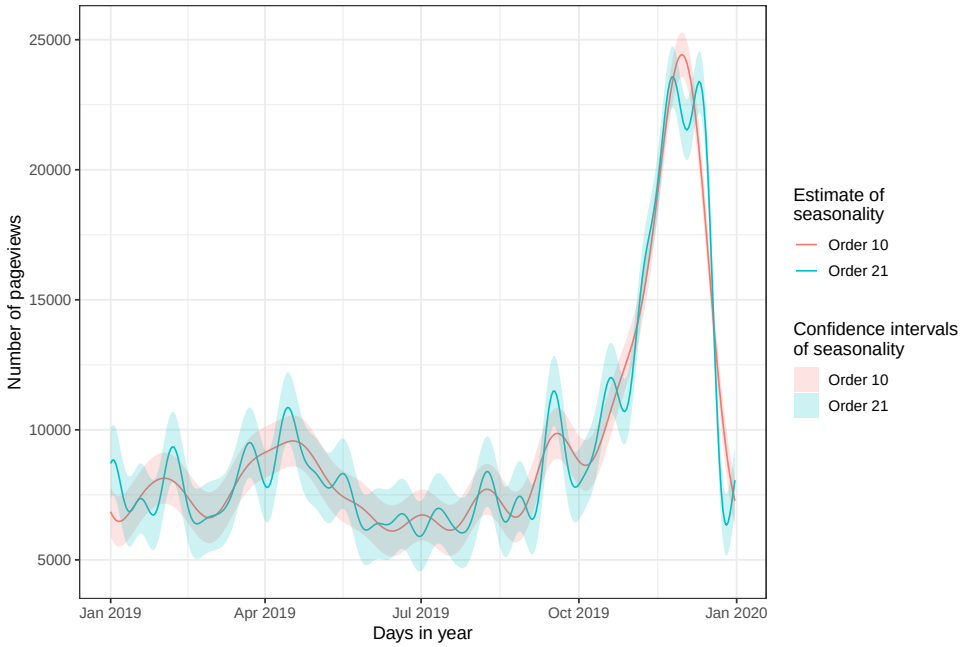


Figure 2 Fourier transformation approximating the yearly seasonal pattern

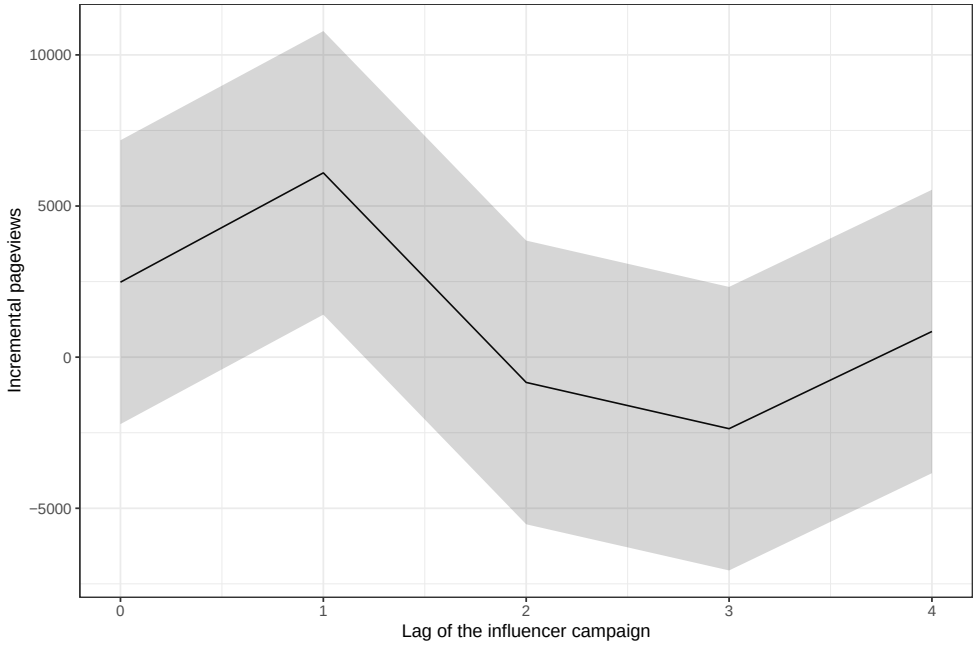


Figure 3 A lag distribution of the influencer campaign – the highest pageviews increase is one day after the campaign.

6 Conclusion

This paper uses the Fourier transformation to analyze seasonal patterns in the study, which also looks at how an influencer marketing effort affects brand awareness. The Akaike information criterion was used to assess the outcomes of a grid search of Fourier transformation order combinations. The ANOVA test revealed that the intra-month pattern was negligible. However, the yearly seasonality was estimated well with the Fourier transformation of order 21. The estimated parameters for the influencer campaign's impact were the study's most remarkable findings; they showed an initial rise in pageviews due to the campaign but a subsequent decline on the third and fourth days. However, the effect was moderate, and the increase in pageviews brought on by the influencer campaign was statistically inconclusive.

7 Discussion

The findings of this study offer insightful information about how influencer marketing affects brand awareness. It is significant to highlight that the study does not consider the true impact on company KPIs like transactions. More research may be done to learn more about the connection between brand awareness and real sales. Furthermore, the study's approach to seasonality decomposition might not be ideal and could be strengthened by using alternative techniques. However, this study offers a helpful framework for additional study in this field.

Funding

This research was supported by the Internal Grant Agency of the Prague University of Economics and Business under grant F4/22/2023.

References

[1] Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*. Volume 33. Issue 1. Pages 178–196. ISSN 0012-9682.

[2] Barth, J. R. & Bennett, J. T. (1974). Cyclical Behavior, Seasonality, and Trend in Economic Time Series. *Nebraska Journal of Economics and Business*. Volume 13. Issue 1. Pages 48-69. ISSN 2327-8234.

[3] Casaló, L., Flavián, C. & Ibáñez-Sánchez, S. (2020). Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*. Volume 117. Pages 510–519. ISSN 0148-2963.

[4] De Veirman, M., Cauberghe, V. & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*. Volume 36. Issue 5. Pages 798–828. ISSN 0265-0487.

[5] Lou, C. & Yuan, S. (2019). Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. *Journal of Interactive Advertising*. Volume 19. Issue 1. Pages 58–73. ISSN 1525-2019.

[6] Nerlove, M., Grether, D. M. & Carvalho, J. L. (1995). *Analysis of Economic Time Series. Economic Theory, Econometrics, and Mathematical Economics.*. Elsevier. ISBN 0-12-515751-7.

[7] Sokolova, K. & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. *Journal of Retailing and Consumer Services*. Volume 53. ISSN 0969-6989.

[8] Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning. ISBN 978-1-111-53439-4.

The Exact Solution of Vehicle Routing Problem by Mixed Integer Linear Programming in Matlab

Jaromír Zahrádka[1]

Abstract. This contribution comes up with a specific solution of the vehicle routing problem. The driver has to deliver the goods from the central warehouse to n customers as efficiently as possible. Each customer has ordered goods that fill a certain number of containers. Each customer point of delivery is given by GPS coordinates. The objective of the solution is to select the number of vehicles and their routes between customers in such a way that the total travel time, including the time for unloading the goods, is as short as possible. Each delivery point is visited only once by one of the vehicles. All used vehicles have a pre-limited capacity of containers. All vehicles return to the central warehouse. In this contribution, the algorithm of the exact solution of the vehicle routing problem was created, which can be used in general for any number n of customers. The algorithm is implemented in Matlab code.

Keywords: Matlab code, mixed integer linear programming, optimization, point of delivery, vehicle routing problem

JEL Classification: C64
AMS Classification: 68W04, 90C11, 05C20

1 The Vehicle Routing Problem

The vehicle routing problem (VRP) is described in [1]. Our solution came from the principles of integer programming and exact algorithms which are listed in [4, 5]. The basic general principles of operating intelligent transport systems published in [2] are respected. The VRP solution concept used in this paper is similar to that of the traveling salesman problem in [6].

1.1 Mathematical Formulation

The vehicle routing problem can be presented as the subsequent graph problem. Let $G = (V, E)$ be a complete directed graph where $V = \{0,1,...,n\}$ is the nodes set and E is the set of all oriented arcs. The truck depot is marked by 0. Nodes $i = 1, 2, ..., n$ correspond to the customers, each with a number q_i of demand containers, which form the row vector $\mathbf{q} = (q_1, q_2, \dots, q_n)$. Each oriented $arc(i, j)$ is associated with non-negative values d_{ij} travel distance (in meters), and c_{ij} travel time (in sec) from node i to node j. For easier references, let $I = \{1, \dots, n\}$, and $I_0 = \{0\} \cup I$. The distance matrix $\mathbf{D} = \left(d_{ij}\right)_{i,j \in I_0}$ and time distance matrix $\mathbf{C} = \left(c_{ij}\right)_{i,j \in I_0}$ are the non-negative and asymmetric.

The VRP consists of finding a collection of k simple cycles, each corresponding to a vehicle route with minimal sum of the distances or time distances of the cycle arcs, such that:

a) each cycle visits the depot - node 0;
b) each vertex $j \in I$ is visited by exactly one cycle;
c) the sum of delivered containers during a cycle does not exceed the vehicle capacity Q.

The order of the customers visited is not limited. For each customer $i \in I$ let m_i be the service time associated with the unloading of goods and dealing with the customer. The service times form a row vector $\mathbf{m} = (m_1, m_2, ..., m_n)$.

[1] University of Pardubice, Department Mathematics and Physics, Studentská 95, 53210 Pardubice, jaromir.zahradka@upce.cz.

1.2 Mathematical Solution

The main method for the exact solution of the vehicle routing problem is optimization implemented by using mixed integer linear programming, which is generally described by

$$\min_{\mathbf{V}} (\mathbf{f}\cdot\mathbf{V}) \; subject\; to \begin{cases} \mathbf{V}_{intcon}\ are\ integers \\ \mathbf{A}\cdot\mathbf{V} \le \mathbf{b} \\ \mathbf{A}_{eq}\cdot\mathbf{V} = \mathbf{b}_{eq} \\ \mathbf{l}_b \le \mathbf{V} \le \mathbf{u}_b \ . \end{cases} \tag{1}$$

The vector $\mathbf{V}$ is the column vector of all flow variables; $\mathbf{f}\cdot\mathbf{V}$ is the objective function with the coefficients contained in the row vector $\mathbf{f}$; $\mathbf{V}_{intcon}$ is the list of variable indices of the vector $\mathbf{V}$ that takes only the integer values; $\mathbf{A}\cdot\mathbf{V}\le\mathbf{b}$ denotes the system of inequality constrains; $\mathbf{A}_{eq}\cdot\mathbf{V}=\mathbf{b}_{eq}$ denotes of the system of linear equations; and $\mathbf{l}_b \le \mathbf{V} \le \mathbf{u}_b$ indicates the lower and upper limits of flow variables.

The linear inequality matrix $\mathbf{A}$ is specified as a matrix of real numbers, which are the linear coefficients in the system of inequality constraints. The right sides of the inequality constraints are included in column vector $\mathbf{b}$. The linear equality constraint matrix $\mathbf{A}_{eq}$ is specified as the matrix of real numbers, which are the linear coefficients in the system of equations. The right sides of the system of equalities are included in column vector $\mathbf{b}_{eq}$.

The lower and upper bounds of flow variables (components of $\mathbf{V}$) are specified as real numbers - elements of column vectors $\mathbf{l}_b$, and $\mathbf{u}_b$.

The core of the practical solution of VRP is to find the appropriate number of k cycles in the graph G which includes all nodes of the graph and which gives the shortest total length of all cycles (or the shortest total driving time of all vehicles). For this purpose, integer flow variables x_{ij} for $i, j \in I_0$ are introduced, which can only take the values 0 or 1 (binary values). The value $x_{ij}=1$ means that the arc from node i to j is included in one cycle and the value $x_{ij}=0$ means that the corresponding arc is not included. For systemic reason variables, x_{ii} are used but all are fixed $x_{ii}=0$, for each $i \in I_0$. Variables x_{ij} are elements of a matrix $\mathbf{X}=\left(x_{ij}\right)_{i,j\in I_0}$, and the number of x_{ij} variables is $(n+1)^2$.

In our work we use other specific integer flow variables, y_i, for each $i \in I$, which indicate the number of containers that were unloaded to customers during the journey from the depot up to and including the node i. The variables y_i are n elements of the vector $\mathbf{y}=(y_1, y_2, \ldots, y_n)$. The sum of all flow variables is $(n+1)^2+n$.

Elements of matrix $\mathbf{X}=\left(x_{ij}\right)_{i,j\in I_0}$ and vector $\mathbf{y}=(y_1, y_2, \ldots, y_n)$ are arranged in row vector $\mathbf{V}$ so that, the transposed vector $\mathbf{V}$ is $\mathbf{V}^T=(\mathbf{X},\mathbf{y})=(x_{00}, x_{01}, \ldots, x_{0n}, x_{10}, x_{11}, \ldots x_{1n}, \ldots, x_{n0}, x_{n1}, \ldots, x_{nn}, y_1, y_2, \ldots, y_n)$.

The solution of VRP is realized like the optimal solution of a mixed-integer linear programming problem:

$$\min_{(\mathbf{X},\mathbf{y})} \left\{ \sum_{i,j=0}^{n} d_{ij}\cdot x_{ij} + \sum_{i=1}^{n} y_i / 100 \right\} \text{ subject to} \tag{2}$$

$$x_{ij},\ i,j \in I_0 \text{ are binary, } x_{ij} \in \{0,1\} \tag{3}$$

$$y_i,\ i \in I \text{ are integer} \tag{4}$$

$$Q\,x_{ij} + y_i - y_j \le Q - q_j, \quad i,j \in I,\ i \ne j \tag{5}$$

$$\sum_{j\in I_0} x_{ij} = 1, \quad i \in I \tag{6}$$

$$\sum_{i\in I_0} x_{ij} = 1, \quad j \in I \tag{7}$$

$$\sum_{j\in I} x_{0j} - \sum_{i\in I} x_{i0} = 0 \tag{8}$$

$$x_{ii} = 0, \quad i \in I_0 \tag{9}$$

$$0 \le x_{ij} \le 1, \quad i, j \in I_0 \tag{10}$$

$$q_j \le y_j \le Q, \quad j \in I \tag{11}$$

In the expressed model (2) is minimized the linear optimization function

$$\mathbf{f} = \sum_{i,j=0}^{n} d_{ij} x_{ij} + \sum_{i=0}^{n} y_i / 100 \,. \tag{12}$$

The main part $\sum_{i,j=0}^{n} d_{ij} x_{ij}$ (the sum of distance in meters) of the optimized function guarantees finding a collection of cycles such that the sum of their lengths is minimal. The second part $\sum_{i=1}^{n} y_i / 100$ of the optimization function (12) is about four orders of magnitude smaller and does not affect the optimization according to the smallest length, but it guarantees that the generated flow values of the y_i is indeed the smallest as possible. Constraint (5) defines $n(n-1)$ conditions between flow variables x_{ij} and numbers y_i and y_j of unloaded containers at nodes i, and j. In the case $x_{ij} = 1$, the inequality (5) expresses the relationship $y_j \ge y_i + q_j$. Thanks to the inclusion of the variables $y_1, y_2, \ldots, y_n$ in the optimized function (12), it is ensured that the values of y_j will always be minimal, i.e. the equation $y_j = y_i + q_j$ will be applied instead of an inequality. In the case $x_{ij} = 0$, for $i \ne j$, the inequality (5) expresses the relationship $y_j \ge y_i + q_j - Q$.

Statements (6) and (7) declare $2n$ equations, which express that only one arc leads from each node $i \in I$ and only one arc leads to each node $j \in I$. Statement (8) declares that the number of arcs leaving the node number 0 is equal to the number of arcs entering the node 0. Statement (9) declares that each $x_{ii} = 0$ (no loops).

The inequalities in (10) declare that the lower and upper bounds of flow variables x_{ij} are 0 and 1. The inequalities in (11) express that the each flow variable y_j is greater than or equal to q_j, and each y_j can't be greater than the capacity Q of vehicle.

1.3 Transformation into Matlab

The distance matrix $\mathbf{D} = \left(d_{ij}\right)_{i,j \in I_0}$ is transformed into Matlab environment as matrix `D`, with the row and column indices *i, j = 1,2, ... ,n+1*. The components `D(i,j)` correspond to the distances $d_{i-1\,j-1}$ of the nodes $i-1$ and $j-1$. Similarly time distance matrix $\mathbf{C} = \left(c_{ij}\right)_{i,j \in I_0}$, is transformed into matrix `C`, i.e. each component `C(i,j)` corresponds to the driving time distance $c_{i-1\,j-1}$.

Variable	n	Q	**A**	**b**	$\mathbf{A}_{eq}$	$\mathbf{b}_{eq}$	$\mathbf{l}_b$	$\mathbf{u}_b$	**m**	**q**	**v**	$\mathbf{V}_{intcon}$	**D**	**C**
Matlab identifier	`n`	`Q`	`A`	`b`	`Aeq`	`beq`	`lb`	`ub`	`m`	`q`	`V`	`intcon`	`D`	`C`

Variable	d_{ij}	c_{ij}	**X**	**X**	x_{ij}
Matlab identifier	`D(i+1,j+1)`	`C(i+1,j+1)`	`X`	`V((1:(n+1)^2,1)`	`V((n+1)*i+j+1,1)`

Variable	**y**	**y**	y_i	$d_{TotLgth}$	t_{TotDur}
Matlab identifier	`y`	`V((n+1)^2+1:(n+1)^2+n,1)`	`V((n+1)^2+i)`	`TotLgth`	`TotDur`

Table 1 The transformations of variables to Matlab identifiers

The created procedure for VRP solving in the Matlab code is included in the M-function `VRP_SOLVER_Za.m` and it is fully listed as an Appendix at the end of the article. The key to transforming used variables into Matlab identifiers can be found in Table 1. The main output variable is the column vector *V* of flow variables, which is obtained

as an output of the command `V=intlinprog(f,intcon,A,b,Aeq,beq,lb,ub,[],options)` (App. row No. 16). A more detailed explanation of command `intlinprog` can be found in the User's Guide [3].

For the solution of VRP via the `intlinprog` command, all flow variables are arranged in a column vector `V` with $(n+1)^2+n$ components. The first $(n+1)^2$ flow variables are integer variables x_{ij}, and each variable x_{ij}, $i, j \in I_0$ is represented by the Matlab flow variable `V(i*(n+1)+j+1,1)`. The last n flow variables of `V` are the values $y_1, y_2, \ldots, y_n$, and each variable y_i, $i \in I$ is represented by `V((n+1)^2+i,1)`.

The objective function of the mixed-integer linear programming problem (12) is, in the Matlab code, expressed like `f'*V`, where `f` is a column vector of coefficients with $(n+1)^2+n$ components. The first $(n+1)^2$ components are elements of the distance matrix `D` so that `f((i-1)*(n+1)+j, 1)=D(i, j)`, $i,j \in \{1,2,\ldots,n+1\}$ (the Appendix, row No. 14).

Alternatively we can use the elements of the time distance matrix `C` so that `f((i-1)*(n+1)+j, 1)=C(i, j)`, $i,j \in \{1,2,\ldots,n+1\}$ (the Appendix, row No. 14, listed as a note after the `%` sign). For the last n components of `f` we use the value 0.01 (the Appendix, row No. 15). It is important to guarantee the optimization of the first $(n+1)^2$ terms. The optimization of the last n-flow variables is of secondary concern.

The vector `intcon` in the command `intlinprog` specifies indices of flow variables, which are taken integers (listend at the end of row No. 15) i.e. command `intcon = 1:(n+1)^2+n`. This means in this case that all flow variables are integers. The constraints (5) give the system of n^2-n linear inequalities with $(n+1)^2+n$ variables. The matrix `A` of system inequalities and the column vector *b* of right sides are created for any n in Matlab code statements on lines No. 3 to 6 in the Appendix. The constraints (6), (7) and (8) give the system of n^2 linear equalities with $(n+1)^2+n$ variables. The matrix `Aeq` of system equalities and the column vector `beq` of right sides are created for any n in the Matlab code statements on lines No. 3 and 7 to 11 of the Appendix.

The other two input variables of the `intlinprog` command (row No. 2 in the Appendix) are the column vectors `lb` and `ub` of lower and upper bounds of the flow variables. With respect to the relations (9), (10), (11) the components of vectors `lb` and `ub` are filled by commands on rows No. 11, 12, 13 in the Appendix. By installation of input variables `f`, `intcon`, `A`, `b`, `Aeq`, `beq`, `lb`, `ub`, and `options` (row No. 2 of the Appendix) in the command `intlinprog`, and running it (the line No. 16), Matlab gives the optimal VRP solution, i.e. the vector of flow variables `V`. The first $(n+1)^2$ components of `V` can be reshaped to square matrix `X` (you can see it on row No. 17 in the Appendix). The last n components of `V` are the components of the vector `y`.

The variables `X(i,j)` which take the value 1, determine the arcs of cycles which make an optimized solution. The two for-cycle commands on lines No. 17, 18, 19, 20 allow calculation of the total length of cycles `TotLgth` and the total driving time `TotDur` of all used vehicles, which is increased by customer service times (command on row No. 17 of the Appendix). The input data for the M-function `SOLVER_VRP_Za.m`, the execution, output variable processing, and drawing output cycles are have to be done using a startup M-script that is not listed in this article.

2 Applied Task

The created M-function `SOLVER_VRP_Za.m` was applied to the delivery of food products from a central warehouse (depot) in the east region of Bohemia to other parts of the Czech Republic. Nineteen customers were

	Depot	Customers								
i	0	1	2	3	4	5	6	7	8	9
E_i (°)	15.90839	16.24160	16.08722	14.69055	15.27025	16.66231	16.98885	14.90985	17.11870	15.58698
N_i (°)	50.02656	50.57513	50.27706	49.41570	49.95616	49.72949	49.95986	50.42491	49.46988	49.61639
q_i	-	7	6	5	4	4	3	3	3	3
	Customers									
i	10	11	12	13	14	15	16	17	18	19
E_i (°)	14.85888	14.74052	15.63217	15.36320	15.598478	15.30779	15.04006	15.22476	16.65998	15.99651
N_i (°)	49.99393	50.15575	50.35722	50.43471	49.42432	50.73916	50.76432	49.42156	49.49256	49.84431
q_i	3	1	2	1	2	1	2	1	1	3

Table 2 The deport and customers GPS coordinates E_i , N_i and numbers of delivered containers q_i

selected for the application task. Each customer ordered a certain number of containers filled with goods. The location of the depot and the location of the customers are indicated using GPS coordinates in Table 2, along with the number of delivered containers. Figure 1 shows the locations of the depot and customers. The distances (in meters) and time distances (in seconds) between the locations of the depot and customers, i.e. the elements of the distance matrix **D** and time distance matrix **C**, were taken from the logistics company operating the transport. They are used in the setup program. Distances and time distances can be realistically obtained based on GPS coordinates using geographic navigation applications specialized for trucks.

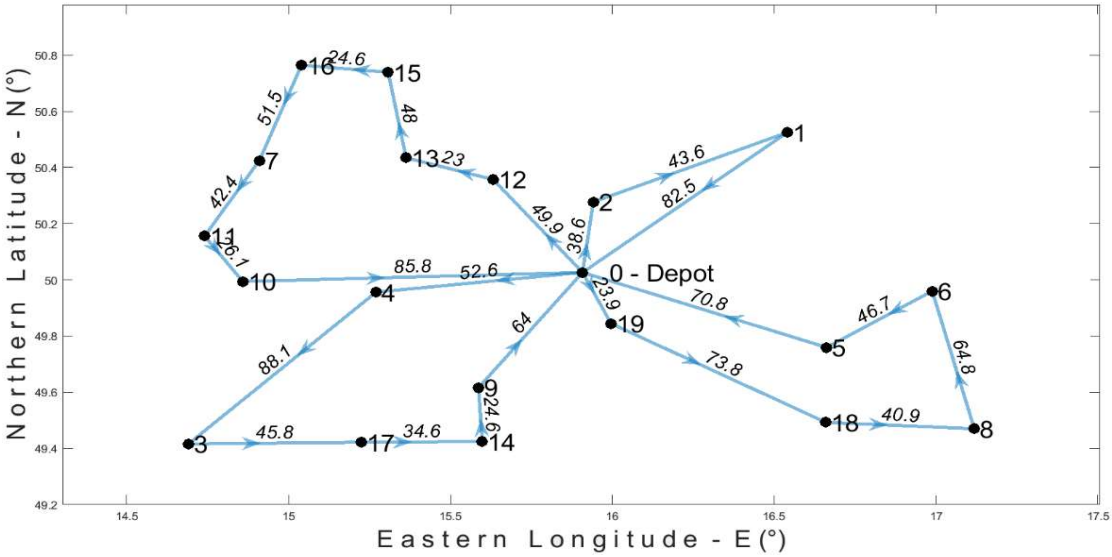


Figure 1 The optimal solution VRP for minimal traveling distance

By running the *VRP_SOLVER_Za.m* function with above-given input parameters, four cycles were found to be the optimal solution. This means that the sum of all four travel distances for trucks is minimal. All four cycles are shown in Figure 1. The list of customers on the cycle route with times of arrival and departure to and from

Cycle 1	**Start**	Customers							Return	Cycle lengths (km)	Duration of Cycles (hh:mm:ss)
i	0	2	1	-	-	-	-	-	0		
q_i	-	6	7	-	-	-	-	-	-		
t_{arr_i}	-	8:00:00	9:22:24	-	-	-	-	-	11:34:32	164.716	4:25:22
t_{dep_i}	7:09:10	8:26:00	9:49:24	-	-	-	-	-	-		
Cycle 2	**Start**	Customers							**Return**		
i	0	4	3	17	14	9	-	-	0		
q_i	-	4	5	1	2	3	-	-	-		
t_{arr_i}	-	8:00:00	10:16:35	11:36:17	12:37:44	13:29:29	-	-	15:12:00	309.575	8:16:02
t_{dep_i}	6:55:58	8:24:00	10:41:35	11:57:17	12:59:44	13:52:29	-	-	-		
Cycle 3	**Start**	Customers							**Return**		
i	0	12	13	15	16	7	11	10	0		
q_i	-	2	1	1	2	3	1	3	-		
t_{arr_i}	-	8:00:00	8:47:14	10:04:12	10:54:39	12:19:57	13:41:02	14:37:21	16:42:36	351.270	9:38:38
t_{dep_i}	7:03:58	8:22:00	9:08:14	10:25:12	11:16:39	12:42:57	14:02:02	15:00:21	-		
Cycle 4	**Start**	Customers							**Return**		
i	0	19	18	8	6	5	-	-	0		
q_i	-	3	1	3	3	4	-	-	-		
t_{arr_i}	-	8:00:00	9:55:50	11:05:07	12:53:25	14:09:34	-	-	15:52:44	351.442	8:25:28
t_{dep_i}	7:27:16	8:23:00	10:16:50	11:28:07	13:16:25	14:33:34	-	-	-		
									Total	$d_{TotLgth}$	t_{TotDur}
										1146.319	30:45:30

Table 3 The customers cycles, arrive and departure times, length and duration of cycles.

customers, numbers of delivered containers, cycle lengths, and duration of cycles, can be found in Table 3. The delivery of goods to customers can be ensured by four trucks, the total distance is 1146.919 km. The sum of driving times (hh:mm:ss) of all four trucks, including the service time for unloading goods, is 30:45:30.

3 Conclusion

The main result of this contribution is the construction of a linear programming model (2) for minimizing travel distance or duration travel time of trucks delivering ordered containers to customers. An integral part is the implementation of the created model in Matlab, i.e. the creation of an M-function *VRP_SOLVER_Za.m* that realizes the solution of vehicle routing problem for any number of n customers. The created function was applied to 19 customers, and the solution took 17 minutes on a common PC. The function is practically usable for up to 30 customers, but the necessary processing time increases. The optimal solution of VRP ensures the shortest travel length or shortest duration of a business trip, and thus the best solution in terms of cost price.

Acknowledgements

The paper was supported by the grant No. CZ.01.1.02/0.0/0.0/21_374/0027244 "Technology development for intelligent traffic flow management - 2nd part - optimization and extension" of Czech Ministry of Industry and Trade.

References

[1] Eksioglu, B., Vural, A.V. & Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering,* 57, 1472-1483.
[2] Jonak, R., Smutný, Z., Simunek, M. & Dolezel, M. (2020). Rout and Travel Time Optimization for Delivery and Utility Services. *Acta Informatica Pragensia,* (2) 9, 200-209.
[3] Math Works. Inc. (2023). *Optimalization Toolbox™. User´s* Guide. Natick.
[4] Winston, W. L. (1994). *Operations Research. Applications and Algorithms.* Duxbury: Duxbury Press.
[5] Toth, P. & Vigo, D. (1998). *Exact algorithms for vehicle routing.* Boston: Kluwer Academic Publisher
[6] Zahrádka, J. (2022) The Traveling Salesman Problem Solution by Mixed Integer Lin. Programming in Matlab Code. *Journal of Applied and Computational Sciences,* 1(1), 45–52. https://doi.org/10.528/zenodo.6880928

Appendix

```
function [X, y, TotLgth, TotDur] = VRP_SOLVER_Za(n, D, C, Q, q, m, PvInd)
options = optimoptions('intlinprog', 'MaxTime', 3600, 'MaxNodes', 3000000);
p=(n+1)*(n+1); A=zeros(n*n-n, p+n); Aeq=zeros(3*n+2,p+n); TotLgth=0;
k=0; for i=1:n for j=1:n if i~=j k=k+1;A(k, p+i)=1;A(k, p+j)=-1; end; end; end
k=0; for i=1:n  for j=1:n  if i~=j k=k+1; A(k, (n+1)*i+1+j)=Q; end; end; end
for i=1:n*n-n  for j=1:n  if A(i, p+j)==-1 b(i, 1)=Q-q(j); end; end; end
for i=1:n  for j=1:n+1 Aeq(i, (n+1)*i+j)=1; end; beq(i, 1)=1; end
for i=1:n  for j=1:n+1 Aeq(n+i, (j-1)*(n+1)+i+1)=1; end; beq(n+i, 1)=1; end
for i=1:n+1  Aeq(i+2*n, (i-1)*(n+1)+i)=1; beq(i+2*n, 1)=0; end
for i=2:n+1  Aeq(3*n+2, i)=1; end
for i=1:n  Aeq(3*n+2, (n+1)*i+1)=-1; end; beq(3*n+2, 1)=0; lb=zeros(p,1); k=0;
for i=1:n+1 for j=1:n+1 k=k+1;if i==j ub(k, 1)=0;else ub(k, 1)=1;end; end; end
for i=1:n  lb(p+i, 1)=q(i); end; for i=1:n ub(p+i, 1)=Q; end
DT=D'; f=DT(:);  %CT=C'; f=CT(:);
f(p+1:p+n)=ones(1, n)/100; intcon=1:(n+1)^2+n;
V=intlinprog(f, intcon, A, b, Aeq, beq, lb, ub, [], options); V=round(V);
X=V(1:p); X=reshape(X, [n+1, n+1]); X=X'; y=V(p+1: end) ; TotDur=sum(m);
  for i=1:(n+1) for j=1:(n+1)
    if X(i, j)==1 TotLgth=TotLgth+D(i, j); TotDur=TotDur+C(i, j); end
  end; end
end
```

www.csov.eu

mme2023.vse.cz

Published by the Czech Society for Operations Research

Winston Churchill Square 1938/4, 130 67 Prague 3, Czechia